

DESY 96 – 244
ITP–UH–23/96



November 1996
hep-th/9611209

SOLITONS, MONOPOLES, AND DUALITY : from Sine–Gordon to Seiberg–Witten ¹

Sergei V. Ketov ²

*Institut für Theoretische Physik, Universität Hannover
Appelstraße 2, 30167 Hannover, Germany
ketov@itp.uni-hannover.de*

Abstract

An elementary introduction into the Seiberg-Witten theory is given. Many efforts are made to get it as pedagogical as possible, and within a reasonable size. The selection of the relevant material is heavily oriented towards graduate students. The basic ideas about solitons, monopoles, supersymmetry and duality are reviewed from first principles, and they are illustrated on the simplest examples. The exact Seiberg-Witten solution to the low-energy effective action of the four-dimensional N=2 supersymmetric pure Yang-Mills theory with the gauge group SU(2) is the main subject of the review. Other gauge groups are also discussed. Some related issues (like adding matter, confinement, string dualities) are outlined.

¹Supported in part by the ‘Deutsche Forschungsgemeinschaft’ and the ‘Volkswagen Stiftung’

² On leave of absence from: High Current Electronics Institute of the Russian Academy of Sciences,
Siberian Branch, Akademichesky 4, Tomsk 634055, Russia

INTRODUCTION

The recent years gave several remarkable achievements in theoretical high energy physics, which constitute a significant progress in understanding the *strongly coupled* supersymmetric gauge theories and their superstring generalizations. It shed new light on some ‘old’ but still ‘hot’ problems, such as confinement, spontaneous symmetry breaking and the role of supersymmetry. The key concepts behind the new developments are (i) *supersymmetry*, (ii) *holomorphicity*, (iii) *duality*, and (iv) *integrability*.

The significance of that new results is so important, that it already changed many traditional ways of thinking about quantum field theory and string theory. At the same time, the precise content of many recent results about duality was not enough appreciated outside of the relatively small community of scientists working in string theory or in a few related areas. Despite of the appearance of several reviews during the recent two years (see, for example, [1, 2, 3, 4, 5, 6, 7], it is still rather difficult for a non-expert to understand the Seiberg-Witten results [8]. It happens, in particular, because some of the available reviews are not elementary enough, while the other do not contain the pre-requisite information. Also, the information required is usually scattered over many sources. It is the purpose of this paper to provide the information which is really needed. I hope that it may be useful for those students who are willing to understand the logic in the beautiful papers of Seiberg and Witten [8]. The uniqueness of their solution was recently proved from first principles [9, 10]. A solid understanding of the well-established facts about the strong-weak coupling duality in the four-dimensional supersymmetric gauge theories may help to enter the more fascinating world of superstring dualities.

The standard example of duality is provided the two-dimensional *Ising model*. It is defined by taking a set of spins σ_i , whose values are restricted to ± 1 , which live on a square two-dimensional lattice with nearest neighbourhood interactions of strength J . The partition function reads

$$Z(K) = \sum_{\sigma} \exp \left(K \sum_{(ij)} \sigma_i \sigma_j \right) ,$$

where the sum (ij) goes over all the nearest neighbours, while the sum on σ goes over all spin configurations, and $K = J/k_B T$. The Ising model is exactly solvable (Onsager), and it exhibits a first-order phase transition to a ferromagnetic state at a critical temperature T_c . It is also known (Kramers and Wannier) that the Ising partition function can be represented in two different ways as a sum over plaquettes

of a lattice. In the first form, the sum goes over plaquettes of the original lattice with the coupling constant K . In the second form, the sum goes over plaquettes of the *dual* square lattice whose vertices are the centers of the faces of the original lattice, with another coupling K^* , where $\sinh 2K^* = (\sinh 2K)^{-1}$. Both formulations are equivalent, but have different coupling constants. There exists a *duality* symmetry which exchanges high temperature or weak coupling ($K \ll 1$) with low temperature or strong coupling ($K^* \gg 1$). Remarkably, the sole existence of the duality symmetry allows one to exactly determine the critical temperature which must occur as the self-dual point where $K = K^*$ or $\sinh(2J/k_B T_c) = 1$. One may view the very existence of the phase transition as a consequence of the fact that the dual weak and strong coupling regimes can be consistently ‘patched’ together. In addition, one can learn about the strong coupling in the Ising model by considering its weakly-coupled dual formulation. Similarly, in the Seiberg-Witten theory, the leading terms in the quantum effective action at any coupling can be obtained from duality, by using the known weak-coupling behaviour together with some additional information, provided by extended supersymmetry, as regards patching together the different regimes. Duality is not the property of the weak-coupling (perturbative) expansion of the quantum theory, but it is the property of the full (exact) theory.

It is usually very difficult to make any exact dynamical statements about non-perturbative phenomena in the ‘realistic’ Standard Model of elementary particles, which is based on the gauge quantum field theory, even if its ($N = 1$) supersymmetric version is considered. It is nevertheless possible to extract some partial information about its non-perturbative behaviour, whose origin can be most clearly seen in the gauge theories with *extended* ($N > 1$) supersymmetry. We are going to start in Part I with the basic facts about monopoles and instantons, which are the main attributes of non-perturbative physics in field theory. Then, we introduce supersymmetry in Part II. The information collected in Parts I and II is necessary for understanding the Seiberg-Witten results, as well as some of their generalizations in the main Part III.

The material appearing in this review is based on my notes collected for the student seminars at the Institute of Theoretical Physics in Hannover during the Spring and Summer 1996. The notes were used for preparing some of my seminar talks at DESY (Hamburg), JINR (Dubna) and Tomsk State University in Russia, all intended for students and based on the already existing literature. The list of references is very far from being complete. Its only purpose is to help the reader to find his own way through the literature (see e.g., refs. [1–10] and references therein for more).

TABLE of CONTENTS

Part I: Basic Examples of Duality in Field Theory	5
1. Sine-Gordon solitons and Thirring model	5
2. Dirac monopole and electro-magnetic duality	9
3. t'Hooft–Polyakov monopole	13
4. Bogomol'nyi-Prasad-Sommerfield limit	16
5. Witten effect and S duality	21
Part II: Introducing Supersymmetry	25
1. Supersymmetry algebras and their representations	25
2. $N = 1$ field theories and superspace	28
3. $N = 2$ super-Yang-Mills theory	31
4. $N = 4$ super-Yang-Mills theory	35
5. Moduli space of the $N = 2$ SYM theory	37
6. $N = 2$ SYM low-energy effective action and renormalization group	40
Part III: Seiberg-Witten theory	45
1. Quantum moduli space in the $SU(2)$ pure $N = 2$ SYM theory	45
2. Duality transformations	50
3. Seiberg-Witten elliptic curve	52
4. Solution to the low-energy effective action	55
5. Other groups, and adding $N = 2$ matter	57
6. Seiberg-Witten version of confinement	65
7. Conclusion	66
References	68

PART I : BASIC EXAMPLES OF DUALITY IN FIELD THEORY

In this introductory part, various aspects of duality in field theory are discussed. We start with the basic explicit example provided by the Sine–Gordon/Thirring models in two spacetime dimensions. Next, the Dirac quantization condition and the t’Hooft–Polyakov monopole in four spacetime dimensions are derived from the first principles. Taken all together, it provides the necessary background for understanding further developments such as the Bogomo’lnyi bound, the BPS states, the Witten effect and S-duality.

1 Sine-Gordon solitons and Thirring model

Let us consider the two-dimensional relativistic field theory characterized by the action

$$I_{\text{SG}}[\phi] = \int d^2x \left[\frac{1}{2} \partial_\mu \phi \partial^\mu \phi + \frac{\alpha}{\beta^2} (\cos \beta \phi - 1) \right] , \quad (1.1)$$

where α and β are constants, $\alpha > 0$. By expanding the potential, one finds that the $\sqrt{\alpha} \equiv m$ plays the role of the mass parameter for the perturbative ‘*meson*’ excitations (after second quantization), whereas $\beta^2 \equiv \lambda/m^2$ acts as the coupling constant. By changing the variables to $\tilde{\phi} = \beta \phi$ and $\tilde{x}^\mu = m x^\mu = m(t, x)$, one can put the equation of motion into the form

$$\square \tilde{\phi} + \sin \tilde{\phi} = 0 , \quad (1.2)$$

known as the *sine-Gordon* equation. This equation enjoys the discrete symmetries $\tilde{\phi} \rightarrow -\tilde{\phi}$ and $\tilde{\phi} \rightarrow \tilde{\phi} + 2\pi$, and it has constant solutions of vanishing energy,

$$\tilde{\phi}_N = 2\pi N , \quad N \in \mathbf{Z} . \quad (1.3)$$

These solutions are not the only classical solutions of finite energy (generically called *solitons*) for the sine-Gordon equation. Since all such solutions at the spacial infinity must approach $\tilde{\phi}_N$, one can associate with each of them the *topological charge*

$$Q = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dx \frac{\partial \tilde{\phi}}{\partial x} = N_1 - N_2 . \quad (1.4)$$

The corresponding *topological* (i.e. non-Noether) current is given by

$$J^\mu = \frac{1}{2\pi} \varepsilon^{\mu\nu} \partial_\nu \tilde{\phi} . \quad (1.5)$$

which is conserved without using any equations of motion. Note that Q does not contain canonical momenta.

The simplest one-soliton ($Q = \pm 1$) solution can be obtained by a Lorentz boost of a static solution with finite energy, the latter being obtained by solving the one-dimensional classical mechanics problem for the potential $-U(\tilde{\phi}) = \cos \tilde{\phi} - 1$:

$$\frac{1}{2}(\tilde{\phi}')^2 = U(\tilde{\phi}) , \quad \text{or} \quad \tilde{x} - \tilde{x}_0 = \pm \int_{\tilde{\phi}(\tilde{x}_0)}^{\tilde{\phi}(\tilde{x})} \frac{d\phi}{2 \sin(\phi/2)} . \quad (1.6)$$

The solution moving with velocity u reads

$$\tilde{\phi}_{\text{S,A}} = \pm 4 \tan^{-1} \left[\exp \left(\frac{\tilde{x} - \tilde{x}_0 - ut}{\sqrt{1 - u^2}} \right) \right] , \quad (1.7)$$

where \pm stands for a soliton or an anti-soliton, with $Q_{\text{S,A}} = \pm 1$, respectively. More complicated multi-soliton solutions for any $Q \in \mathbf{Z}$ comprising any number of solitons and anti-solitons under collision are also known to exist, and each of them is reducible at $t \rightarrow \pm\infty$ to the sum of the well-separated solitons and anti-solitons up to certain time delays, with the velocities and energy profiles being *unchanged* [11]. It is also clear that a multi-soliton solution with a given Q cannot ‘decay’ into solitons with a different Q because of the topological charge conservation (the *superselection* rule).

The fact that solitons maintain their shape despite their collisions and have finite classical mass (defined by the static energy) suggests their physical interpretation as classical particles. Therefore, we have two apparently different ‘sorts’ of particles in the sine-Gordon theory: the perturbative ‘mesons’ as the small excitations of the second-quantized field with the mass m , and the non-perturbative solitons of mass $M = 8m/\beta^2 = 8m^3/\lambda$, which are the extended classical solutions of the non-linear field equations. The solitons interpolate between different minima of the potential, and they are absent in the perturbative spectrum. In the weak coupling limit, $\lambda \rightarrow 0$ or $\beta \rightarrow 0$, the ‘meson’ mass m is constant or small, while the soliton mass M is large.

In fact, these two ‘sorts’ of particles can be considered on equal footing in the full quantum theory [12]. The whole point is the known *quantum equivalence* of the sine-Gordon model to the *Thirring* model to be defined by the action

$$I_{\text{T}}[\psi] = \int d^2x \left[\bar{\psi} i \gamma^\mu \partial_\mu \psi - m_{\text{F}} \bar{\psi} \psi - \frac{g}{2} (\bar{\psi} \gamma^\mu \psi)^2 \right] . \quad (1.8)$$

The equivalence is established via bosonization [13, 14]:

$$\psi_\pm(x) = \exp \left\{ \frac{2\pi}{i\beta} \int_{-\infty}^x \frac{\partial \phi(x')}{\partial t} dx' \mp \frac{i\beta}{2} \phi(x) \right\} , \quad (1.9)$$

where the two spinor components are distinguished by \pm . One can show that the ψ_{\pm} satisfies the Thirring equations of motion provided the ϕ satisfies the sine-Gordon equation and vice versa. The *vertex operator construction* of eq. (1.9) establishes the equivalence of the correlation functions in both theories, while the correspondence between their coupling constants turns out to be [13, 14]

$$\frac{\beta^2}{4\pi} = \frac{1}{1 + g/\pi} . \quad (1.10)$$

The strong coupling in the T-theory (large g) is thus mapped to the weak coupling (small β) in the *dual* SG-theory and vice versa. It allows one to identify the particles corresponding to the fluctuations of ψ_{\pm} with solitons and anti-solitons. One can show that the meson SG states correspond to the fermion-antifermion *bound* states in the Thirring theory [11].

Actually, we were rather sloppy above, since we ignored the effects of renormalization in quantum field theory. Fortunately, the renormalization effects in the SG- and T-theories are under control, and they can be fully taken into account by normal ordering, in terms of bare parameters m_0^2 and m_F , and the fermionic field renormalization parameters C_{\pm} . One uses the Baker-Hausdorff identity to show that [15]

$$\frac{m^4}{\lambda} : \left[\cos \left(\frac{\sqrt{\lambda}}{m} \phi \right) - 1 \right] := m_0^2 \frac{m^2}{\lambda} \left[\cos \left(\frac{\sqrt{\lambda}}{m} \phi \right) - 1 \right] . \quad (1.11)$$

The action-angle variables, in which the classical SG hamiltonian reduces to a free particle form, are known [16], which implies that the SG model is exactly solvable both as a classical theory and as a quantum one (semiclassical quantization is exact in this case). Accordingly, the quantum renormalization in the SG theory amounts to replacing the naive coupling constant β^2 to a renormalized coupling constant γ ,

$$\gamma = \frac{\beta^2}{1 - \beta^2/8\pi} = \frac{8\pi}{1 + 2g/\pi} . \quad (1.12)$$

The quantum bosonization rules are given by the normal-ordered equation (1.9):

$$\psi_a(x) = C_a : \exp[A_a(x)] : , \quad A_{\pm}(x) = \frac{2\pi m}{i\sqrt{\lambda}} \left(\int_{-\infty}^x \dot{\phi}(x') dx' \right) \mp \frac{i\sqrt{\lambda}}{2m} \phi(x) , \quad (1.13)$$

where $a = \{\pm\} = (1, 2)$. In particular, it implies the relations [13, 14]

$$\begin{aligned} m_0^2(m^2/\lambda) \cos \left[(\sqrt{\lambda}/m) \phi \right] &= -m_F \bar{\psi} \psi , \\ -(\sqrt{\lambda}/2\pi m) \varepsilon^{\mu\nu} \partial_{\nu} \phi &= \bar{\psi} \gamma^{\mu} \psi , \end{aligned} \quad (1.14)$$

while the fermionic charge can be identified with the topological charge.

It is not difficult to show that the fermions defined by eq. (1.13) do satisfy the local Fermi rules [14]. The canonical equal-time commutation relations

$$[\phi(x), \phi(y)]_- = [\dot{\phi}(x), \dot{\phi}(y)]_- = 0, \quad [\phi(x), \dot{\phi}(y)]_- = i\delta(x-y), \quad (1.15)$$

imply that $[A_a(x), A_b(y)]$ is either $i\pi$ or $-i\pi$ when $x \neq y$, which leads to ³

$$\{\psi_a(x), \psi_b(y)\}_+ = 0, \quad \text{and} \quad \{\psi_a(x), \psi_b^\dagger(y)\}_+ = Z\delta(x-y)\delta_{ab}, \quad (1.16)$$

where Z is another renormalization constant. In addition one finds that [11]

$$[\phi(y), \psi(x)] = (2\pi/\beta)\theta(x-y)\psi(x), \quad x \neq y. \quad (1.17)$$

Being applied to the soliton state with $\phi(\infty) - \phi(-\infty) = 2\pi/\beta$, the operator ψ thus reduces it to a state in the vacuum sector with $\phi(\infty) - \phi(-\infty) = 0$. Because of eq. (1.17), $\psi(x)$ alters a field ϕ by a step function which can be considered as a ‘point soliton’ (obviously, a local operator cannot create an extended object). The physical (extended) soliton then arises via interactions. The ψ and ψ^\dagger can therefore be interpreted as the destruction and creation operators for bare solitons.

One learns from the explicit duality between the T- and SG-models that

- duality is a quantum correspondence which relates the strong coupling in one theory with the weak coupling in another theory;
- duality interchanges ‘fundamental’ quanta with solitons, and thus establishes a ‘democracy’ between them;
- in addition, duality exchanges Noether currents with topological currents.

In other words, the full physical spectrum does not only contain the particles corresponding to the fields present in the classical Lagrangian, but it also contains other particles which correspond to the soliton solutions and which are required by duality.

It is highly non-trivial to generalize that ideas to *four* dimensions. In particular, the naive generalization of the two-dimensional sine-Gordon theory to a scalar field theory in higher dimensions does not work. ⁴ Hence, the need for some additional *gauge* fields becomes apparent. Moreover, we need a gauge theory in which the semiclassical properties are not destroyed by quantum corrections. It is the (extended) *supersymmetric* gauge theories that enjoy such a behaviour. In what follows, both ideas will be discussed in some detail.

³The renormalization coefficients C_\pm are to be adjusted in the coincidence limit.

⁴The absence of non-trivial static solutions for a very general class of scalar potentials in more than two dimensions is known as the *Derrick* theorem [17].

2 Dirac monopole and electro-magnetic duality

The Maxwell equations for the electromagnetism in 1+3 dimensions can be written down in the relativistic form as

$$\partial_\nu F^{\mu\nu} = -j_e^\mu, \quad \partial_\nu {}^*F^{\mu\nu} = 0, \quad (2.1)$$

or in the vector form as

$$\begin{aligned} \operatorname{div} \vec{E} &= \rho_e, & \operatorname{rot} \vec{E} + \frac{\partial \vec{B}}{\partial t} &= 0, \\ \operatorname{div} \vec{B} &= 0, & \operatorname{rot} \vec{B} - \frac{\partial \vec{E}}{\partial t} &= \vec{J}_e, \end{aligned} \quad (2.2)$$

where we use the notation $\mu = (0, i) = (0, 1, 2, 3)$, $\eta_{\mu\nu} = \operatorname{diag}(-, +, +, +)$, $\varepsilon^{0123} = 1$, $j_e^\mu = (\rho_e, \vec{J}_e)$, and define ⁵

$$F^{0i} = -E^i, \quad F^{ij} = -\varepsilon^{ijk} B^k, \quad \text{and} \quad {}^*F^{\mu\nu} = \frac{1}{2} \varepsilon^{\mu\nu\lambda\rho} F_{\lambda\rho}. \quad (2.3)$$

In vacuum, where $\rho_e = \vec{J}_e = 0$, eq. (2.2) can be rewritten to the form

$$\vec{\nabla} \cdot (\vec{E} + i\vec{B}) = 0, \quad \vec{\nabla} \wedge (\vec{E} + i\vec{B}) = i \frac{\partial}{\partial t} (\vec{E} + i\vec{B}), \quad (2.4)$$

which is invariant under the duality rotations ⁶

$$\vec{E} + i\vec{B} \rightarrow e^{-i\theta} (\vec{E} + i\vec{B}), \quad (2.5)$$

parametrized by an arbitrary angle θ . In particular, when $\theta = \pi/2$, there is a discrete symmetry

$$D: \quad \vec{E} \rightarrow +\vec{B}, \quad \vec{B} \rightarrow -\vec{E}, \quad (2.6)$$

whose square $D^2: (\vec{E}, \vec{B}) \rightarrow (-\vec{E}, -\vec{B})$ is just the charge conjugation C . Eq. (2.6) is obviously equivalent to

$$F^{\mu\nu} \rightarrow {}^*F^{\mu\nu}, \quad {}^*F^{\mu\nu} \rightarrow -F^{\mu\nu}, \quad (2.7)$$

and it can only be valid in 1+3 dimensions because of the identity $(*)^2 = -\mathbf{1}$. ⁷

⁵We normally take $c = 1$ and $\hbar = 1$, but sometimes reintroduce one or both of them, in order to emphasize the relativistic and/or quantum nature of some equations.

⁶The Maxwell equations in vacuum are also known to be invariant under Lorentz and conformal transformations.

⁷Only in 1+3 dimensions do the electric and magnetic fields both constitute vectors.

The energy and momentum density of the electro-magnetic field,

$$\frac{1}{2} |\vec{E} + i\vec{B}|^2 = \frac{1}{2} (\vec{E}^2 + \vec{B}^2) , \quad (2.8)$$

and

$$\frac{1}{2i} (\vec{E} + i\vec{B})^* \wedge (\vec{E} + i\vec{B}) = \vec{E} \wedge \vec{B} , \quad (2.9)$$

respectively, are invariant under the duality (2.5). As far as the Lagrangian and the topological charge density are concerned, they are given by the real and imaginary part of

$$\frac{1}{2} (\vec{E} + i\vec{B})^2 = \frac{1}{2} (\vec{E}^2 - \vec{B}^2) + i\vec{E} \cdot \vec{B} , \quad (2.10)$$

respectively, and, hence, they transform as a doublet under the duality [1].

The duality symmetry is lost if an electric current j^μ enters the Maxwell equations. Therefore, if we want to keep the electro-magnetic duality in the presence of matter, we have to add magnetic source terms into the Maxwell equations as well, so that

$$\partial_\nu {}^*F^{\mu\nu} = -k^\mu \neq 0 . \quad (2.11)$$

For example, the discrete duality transformations (2.7) are to be appended by

$$j^\mu \rightarrow k^\mu , \quad \text{and} \quad k^\mu \rightarrow -j^\mu . \quad (2.12)$$

If the duality makes sense, it has also to be consistent with quantum mechanics and non-abelian gauge theories (see also the next section). Consider a charged quantum particle with momentum \vec{p} , whose interaction with the electromagnetic field via the standard substitution $\vec{p} = -i\vec{\nabla} \rightarrow -i(\vec{\nabla} - ie\vec{A})$ is governed by a potential $A^\mu = (A_0, \vec{A})$ to be defined from the field strength

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu . \quad (2.13)$$

The Schrödinger equation for the quantum particle,

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2m}(\vec{\nabla} - ie\vec{A})^2\psi + V\psi , \quad (2.14)$$

is invariant under the gauge transformations

$$\psi \rightarrow e^{-ie\chi}\psi , \quad \vec{A} \rightarrow \vec{A} - \vec{\nabla}\chi = \vec{A} - \frac{i}{e}e^{ie\chi}\vec{\nabla}e^{-ie\chi} , \quad (2.15)$$

where the gauge parameter χ enters via the $U(1)$ group element $e^{ie\chi}$, which must be single-valued and continuous. However, it is the potential A^μ itself that gives a problem since its definition (2.13) apparently implies that $\partial_\nu {}^*F^{\mu\nu} = -k^\mu = 0$.

Therefore, the electromagnetic potential of a magnetic charge (called *monopole*), if exists, has to be singular inside the monopole.⁸ The consistent solution outside the monopole of magnetic charge g , resulting in a magnetic field

$$\vec{B} = \frac{g\vec{e}_r}{4\pi r^2} , \quad (2.16)$$

makes use of the ambiguity relating the vector potential to the field strength [18]: one can use different potentials in different regions if their differences in the overlapping regions are given by gauge transformations. It is the physically measurable field strength $F^{\mu\nu}$ that has to be continuous and unambiguous. The simplest way out is to divide a sphere S^2 surrounding the monopole into a northern (N) and southern (S) hemispheres, corresponding to $0 \leq \theta \leq \pi/2$ and $\pi/2 \leq \theta \leq \pi$, respectively, the equator (E) with $\theta = \pi/2$ being the overlap region. A non-singular solution to the vector potential on the hemispheres reads⁹

$$\begin{aligned} \vec{A}_N &= + \frac{g}{4\pi r} \frac{1 - \cos \theta}{\sin \theta} \vec{e}_\phi , \\ \vec{A}_S &= - \frac{g}{4\pi r} \frac{1 + \cos \theta}{\sin \theta} \vec{e}_\phi , \end{aligned} \quad (2.17)$$

so that $\vec{B} = \vec{\nabla} \times \vec{A}$ just yields eq. (2.16). This construction makes sense, since the difference of the vector potentials at $\theta = \pi/2$,

$$\vec{A}_N - \vec{A}_S = -\vec{\nabla}\chi , \quad \chi = -\frac{g}{2\pi}\phi , \quad (2.18)$$

is a gauge transformation indeed, while the enclosed magnetic charge is given by

$$g = \int_{S^2} \vec{B} \cdot d\vec{S} = \int_N \vec{B} \cdot d\vec{S} + \int_S \vec{B} \cdot d\vec{S} = \int_E (\vec{A}_N - \vec{A}_S) \cdot d\vec{l} = \chi(0) - \chi(2\pi) \neq 0 , \quad (2.19)$$

as required. The gauge transformation parameter χ in eq. (2.18) is not a continuous function, but it is the function $e^{-ie\chi}$ that has to be continuous so that $\exp(-ieg) = 1$. Reintroducing \hbar and c , one can represent it in the form

$$eg = 2\pi n\hbar c , \quad n \in \mathbf{Z} , \quad (2.20)$$

known as the celebrated *Dirac quantization condition* [19].

In mathematical terms, the sphere S^2 surrounding the monopole is just the base space of a non-trivial $U(1)$ principal fibre bundle. The resulting structure is a manifold when the fibers are patched together in a globally consistent way, with gauge

⁸Since we do not expect the electrodynamics to be a correct theory at very small distances, the existence of singularity at the location of a monopole does not pose a serious problem.

⁹A general solution can be understood in more abstract terms (see below).

transformations as the transition functions. Because of eqs. (2.18) and (2.19), the magnetic charge of the monopole can be directly interpreted as the *winding number* of the gauge transformation, defining a map from the overlap region (equator) S^1 to the gauge group $U(1) \sim S^1$. These maps are classified by the first *homotopy* group $\pi_1(U(1)) \sim \mathbf{Z}$, whose elements can be identified with the integers n appearing in the Dirac quantization condition (2.20).¹⁰ The same integer is given by the *first Chern class* c_1 of the bundle, which is defined by an integration of the two-form $\frac{1}{2\pi}F_{\mu\nu}dx^\mu \wedge dx^\nu$ over S^2 .

It is clear from eq. (2.20) that just assuming the existence of a monopole¹¹ is sufficient for explaining the quantization of the electric charge e , as well as another well-known experimental fact that the absolute values of the electron and proton electric charges are *exactly* equal. It is also clear from eqs. (2.5) and (2.12) that the electro-magnetic duality requires the rotation of electric and magnetic charges of point particles representing matter, in order to keep the Maxwell equations invariant,

$$e + ig \rightarrow e^{-i\theta}(e + ig) . \quad (2.21)$$

It should be noticed that the Dirac quantization condition (2.20) does *not* respect the symmetry (2.21). It is related to the (unjustified) hidden assumption that the Dirac monopole does not carry an electric charge. In order to generalize eq. (2.20) to the form which is consistent with the electromagnetic duality, one first notices that eq. (2.20) can be obtained in many different ways. For example, when computing the orbital angular momentum

$$\vec{L} = \int d^3r \vec{r} \times (\vec{E} \times \vec{B}) \quad (2.22)$$

of a point particle with an electric charge e in the field of the magnetic monopole with a magnetic charge g , just demanding the \vec{L} be quantized in units of $\hbar/2$ also yields eq. (2.20). Eq. (2.22) can be easily generalized to the case of two *dyons*, having both electric and magnetic charges, (q_1, g_1) and (q_2, g_2) . The momentum quantization then gives rise to the so-called *Dirac-Zwanziger-Schwinger* (DZS) quantization condition [19, 20, 21],

$$q_1 g_2 - q_2 g_1 = 2\pi n, \quad n \in \mathbf{Z} , \quad (2.23)$$

which is invariant under the electromagnetic duality (2.21). The DZS condition implies that the allowed electric and magnetic charges of a dyon are quantized, and they should lie on a two-dimensional lattice [7].

¹⁰In eq. (2.17) above, the case of $n = 1$ was considered.

¹¹No monopoles were observed in the experiments, which implies that, if they nevertheless exist, their masses are to be high enough.

Similarly to the SG–T duality considered in the preceding section, the interchange of electricity and magnetism by exchanging the coupling constants leads to the interchange of weak and strong coupling. Like solitons in the SG theory, the Dirac monopole does not exist in the spectrum of standard quantum electrodynamics, and no local theory exists which could accomodate both electrons and Dirac monopoles.

One learns from the electromagnetic duality that

- it requires magnetic monopoles,
- the existence of monopoles in a gauge theory is closely related to the existence of a compact $U(1)$ gauge group,
- the magnetic charge is given by the topological quantity – the winding number – which belongs to the first homotopy group of $U(1)$,
- electro-magnetic duality implies C -invariance,
- the electric and magnetic charges of dyons lie on a two-dimensional lattice.

The derivation of the Dirac quantization condition above considers a monopole from a distance, so it directly applies to an electron which is not confined unlike the quarks. It is also very general, since no particular underlying theory was used for describing monopoles. However, in order to probe the monopole inside, one needs a deeper gauge theory, which contains both electrically and magnetically charged particles. The so-called *Georgi-Glashow model* is such a theory, as was independently found by t’Hooft and Polyakov [22, 23]. This model is considered in the next section.

3 t’Hooft-Polyakov monopole

The basic idea is to embed the $U(1)$ generator Q of electric charge into a larger compact gauge group, say, $SU(2)$ or $SO(3)$ for simplicity, i.e. to switch to a non-abelian gauge theory. The standard Higgs mechanism can then be used to select the direction of Q amongst the $SO(3)$ generators. The situation is very much analogous to the SG theory (sect.1) having the discrete vacuum symmetry (1.3) which is now replaced by the continuous gauge symmetry.

The Georgi-Glashow model consists of an $SO(3)$ gauge field A_μ^a and a Higgs triplet field Φ^a , with the Lagrangian

$$\mathcal{L}_{\text{GG}} = -\frac{1}{4}F_{\mu\nu}^a F^{a\mu\nu} + \frac{1}{2}D^\mu\Phi^a D_\mu\Phi^a - V(\Phi) , \quad (3.1)$$

where the Yang-Mills field strength

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + e\varepsilon^{abc} A_\mu^b A_\nu^c , \quad (3.2)$$

the covariant derivative

$$D_\mu \Phi^a = \partial_\mu \Phi^a + e\varepsilon^{abc} A_\mu^b \Phi^c , \quad (3.3)$$

and the Higgs potential

$$V(\Phi) = \frac{\lambda}{4} (\Phi^a \Phi^a - v^2)^2 , \quad (3.4)$$

have been introduced. The corresponding equations of motion read

$$D_\mu F^{a\mu\nu} = e\varepsilon^{abc} \Phi^b D^\nu \Phi^c , \quad (D^\mu D_\mu \Phi)^a = -\lambda \Phi^a (\Phi^b \Phi^b - v^2) , \quad (3.5)$$

and the Bianchi identity is

$$D_\mu {}^* F^{a\mu\nu} = 0 . \quad (3.6)$$

Like in the SG theory, our strategy is to find static classical solutions of the equations of motion with a finite energy. The *improved*¹² stress-energy tensor is given by

$$\vartheta^{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta I_{\text{GG}}}{\delta g_{\mu\nu}} = -F^{a\mu\rho} F_\rho^{a\nu} + D^\mu \Phi^a D^\nu \Phi^a - \eta^{\mu\nu} \mathcal{L}_{\text{GG}} , \quad (3.7)$$

and it is classically conformally invariant, $\vartheta_\mu^\mu = 0$, if $\lambda = 0$. It still makes sense to choose $V(\Phi) = 0$ while maintaining $\langle \Phi \rangle \neq 0$ which spontaneously breaks both the gauge and scale invariances, and it is going to be used later, in the next section. Because of eq. (3.7), the energy density reads

$$\vartheta_{00} = \frac{1}{2} (\vec{E}^a \vec{E}^a + \vec{B}^a \vec{B}^a + \Pi^a \Pi^a + \vec{D} \Phi^a \cdot \vec{D} \Phi^a) + V(\Phi) , \quad (3.8)$$

where we have introduced the momenta $\Pi^a \equiv D_0 \Phi^a$, and defined $E^{ai} = -F^{a0i}$, $B^{ai} = -\frac{1}{2} \varepsilon^{ijk} F_{jk}^a$. Obviously, we have $\vartheta_{00} \geq 0$, while $\vartheta_{00} = 0$ if and only if $F^{a\mu\nu} = D^\mu \Phi^a = V(\Phi) = 0$. The Higgs vacuum \mathcal{M}_{H} is therefore given by the vanishing gauge field and a constant Higgs field, $\Phi^a \Phi^a = v^2$, i.e. $\mathcal{M}_{\text{H}} = S^2$. The perturbative spectrum consists of a massless ‘photon’, massive spin-one gauge bosons W^\pm of mass $|e|v$ and a Higgs field whose mass is $v\sqrt{2\lambda}$.

The finite energy solutions must lie in \mathcal{M}_{H} at the spacial infinity, whereas the Higgs field overthere provides a map from S_∞^2 to $\mathcal{M}_{\text{H}} = S^2$. Such maps are topologically classified by the integer winding number which is an element of the second homotopy group of S^2 ,

$$\pi_2(S^2) = \mathbf{Z} . \quad (3.9)$$

¹²The improved stress-energy tensor is symmetric and gauge-invariant by definition.

It is easy to check that finite-energy field configurations with a non-trivial winding number require a non-vanishing gauge field. Indeed, it follows from the relations

$$\vartheta_{00} \geq \int d^3x \frac{1}{2} \vec{\nabla} \Phi^a \vec{\nabla} \Phi^a , \quad (3.10)$$

and

$$(\vec{\nabla} \Phi^a)^2 = \left(\frac{\partial \Phi^a}{\partial r} \right)^2 + (\vec{e}_r \times \vec{\nabla} \Phi^a)^2 , \quad (3.11)$$

that at $A^\mu = 0$ one arrives at a linearly divergent integral,

$$\vartheta_{00} \geq \int_0^\infty \frac{r^2 dr}{r^2} , \quad (3.12)$$

since the non-trivial winding number implies non-vanishing angular derivatives of Φ^a at spacial infinity, and their contribution alone in eq. (3.11) leads to the divergence (3.12) in eq. (3.10). Simultaneously, this argument shows that, in order to achieve a finite-energy solution, there should be a cancellation between the angular part of the vector potential (which must fall as $1/r$) and the angular derivative of Φ , such that the *covariant* derivative of the Higgs field vanishes at spacial infinity.

It is not difficult to see that the $1/r$ falloff in the angular component of the gauge field A_μ^a gives rise to a non-vanishing magnetic field at spacial infinity, i.e. it gives a monopole ! When taking into account only the leading $1/r$ -terms, the general solution to the equation

$$D_\mu \Phi^a = \partial_\mu \Phi^a + e \varepsilon^{abc} A_\mu^b \Phi^c \sim 0 \quad (3.13)$$

reads

$$A_\mu^a \sim - \frac{1}{ev^2} \varepsilon^{abc} \Phi^b \partial_\mu \Phi^c + \frac{1}{v} \Phi^a A_\mu , \quad (3.14)$$

where A_μ is an arbitrary field. Accordingly, the field strength takes the form

$$F^{a\mu\nu} = \frac{1}{v} \Phi^a F^{\mu\nu} , \quad \text{where} \quad F_{\mu\nu} = - \frac{1}{ev^3} \varepsilon^{abc} \Phi^a \partial_\mu \Phi^b \partial_\nu \Phi^c + \partial_\mu A_\nu - \partial_\nu A_\mu . \quad (3.15)$$

The equations of motion (3.5) together with the Bianchi identity (3.6) imply in addition that

$$\partial_\mu F^{\mu\nu} = \partial_\mu {}^* F^{\mu\nu} = 0 , \quad (3.16)$$

outside of the core of the monopole. It is therefore the Higgs field that is solely responsible for the non-vanishing magnetic charge of the gauge field configuration (3.15):

$$g = \int_{S_\infty^2} \vec{B} \cdot d\vec{S} = \frac{1}{2ev^3} \int_{S_\infty^2} dS^i \varepsilon^{ijk} \varepsilon^{abc} \Phi^a \partial^j \Phi^b \partial^k \Phi^c = \frac{4\pi}{e} n , \quad n \in \mathbf{Z} . \quad (3.17)$$

The just found quantization condition,

$$eg = 4\pi n , \quad (3.18)$$

differs by a factor of 2 from the Dirac quantization condition (2.20). It is related to the fact that we could add into our theory more fields in the fundamental representation $\underline{2}$ of $SU(2)$ whose quanta carry an electric charge $\pm e/2$. It is the Dirac quantization condition with respect to the electric charge $\pm e/2$ that yields eq. (3.18).

The main lesson one learns from this section is that there exists a deep connection between the Dirac monopoles and the Higgs mechanism [24], namely,

- finite-energy solutions with non-vanishing topological charge in the Georgi-Glashow model are necessarily magnetic monopoles which satisfy the Dirac quantization condition.

In mathematical terms, on the one hand, given a gauge (simply connected) group G broken down to a subgroup H by the non-vanishing Higgs field vacuum expectation value, the topology of the Higgs vacuum is classified by $\pi_2(G/H)$. On the other hand, the general Dirac monopole configurations to be constructed by patching together H -gauge fields along the equator are classified by $\pi_1(H)$. It is just the topology theorem that

$$\pi_2(G/H) = \pi_1(H) . \quad (3.19)$$

The exact solution to the Georgi-Glashow model in the limit of vanishing potential ($V = 0$) is discussed in the next section.

4 Bogomol'nyi-Prasad-Sommerfield limit

An exact monopole solution with a non-vanishing topological charge, $n \neq 0$, cannot be invariant under the rotational subgroup $SO(3)_R$ of the Lorentz group, because the Higgs fields must vary at spacial infinity. The solution cannot be invariant under the global gauge transformations $SO(3)_G$ either since, otherwise, the Higgs fields must vanish. However, the lowest-energy monopole solution may still be invariant under the *diagonal* subgroup $SO(3)$ of the $SO(3)_R \otimes SO(3)_G$. When imposing this symmetry, one is left with the unique *Ansatz* ¹³

$$\Phi^a = \frac{e_r^a}{er} H(evr) , \quad A_0^a = 0 , \quad A_i^a = -\varepsilon^a_{ij} \frac{e_r^j}{er} [1 - K(evr)] , \quad (4.1)$$

¹³Strictly speaking, the additional discrete symmetry which is a combination of parity and a sign change of Φ has to be imposed too.

in terms of two radial real functions, H and K , subject to the boundary conditions (sect. 3)

$$\begin{aligned} K(\xi) &\rightarrow 1, \quad H(\xi) \rightarrow 0, \quad \text{at} \quad \xi \rightarrow 0, \\ K(\xi) &\rightarrow 0, \quad \frac{H(\xi)}{\xi} \rightarrow 1, \quad \text{at} \quad \xi \rightarrow \infty, \end{aligned} \quad (4.2)$$

where the dimensionless parameter $\xi = evr$ has been introduced. The mass of this static field configuration is determined by eq. (3.8),

$$\begin{aligned} M_M = & \frac{4\pi v}{e} \int_0^\infty \frac{d\xi}{\xi^2} \left[\xi^2 \left(\frac{dK}{d\xi} \right)^2 + K^2 H^2 + \frac{1}{2} \left(\xi \frac{dH}{d\xi} - H \right)^2 \right. \\ & \left. + \frac{1}{2} (K^2 - 1)^2 + \frac{\lambda}{4e^2} (H^2 - \xi^2)^2 \right], \end{aligned} \quad (4.3)$$

whereas the equations of motion (3.5) take the form

$$\begin{aligned} \xi^2 \frac{d^2 K}{d\xi^2} &= KH^2 + K(K^2 - 1), \\ \xi^2 \frac{d^2 H}{d\xi^2} &= 2K^2 H + \frac{\lambda}{e^2} H(H^2 - \xi^2). \end{aligned} \quad (4.4)$$

The system of non-linear differential equations (4.4) for the unknown radial functions H and K admits a finite energy solution, and it can be explicitly integrated in a certain limit [25]. In order to understand the nature of this limit, let us discuss first the so-called *Bogomol'nyi bound* [26]. This bound can be obtained by considering the mass M_M of a static configuration with vanishing electric field,

$$\begin{aligned} M_M &= \int d^3r \left[\frac{1}{2} (\vec{B}^a \cdot \vec{B}^a + \vec{D}\Phi^a \cdot \vec{D}\Phi^a) + V(\Phi) \right] \geq \int d^3r \frac{1}{2} (\vec{B}^a \cdot \vec{B}^a + \vec{D}\Phi^a \cdot \vec{D}\Phi^a) \\ &= \frac{1}{2} \int d^3r (\vec{B}^a - \vec{D}\Phi^a)(\vec{B}^a + \vec{D}\Phi^a) + vg, \end{aligned} \quad (4.5)$$

where we have used the expression for the magnetic charge g in the form

$$g = \int_{S_\infty^2} \vec{B} \cdot d\vec{S} = \frac{1}{v} \int_{S_\infty^2} \Phi^a \vec{B}^a \cdot d\vec{S} = \frac{1}{v} \int \vec{B}^a (\vec{D}\Phi)^a d^3r, \quad (4.6)$$

because of eq. (3.15) and the Bianchi identity $\vec{D} \cdot \vec{B}^a = 0$. Eq. (4.5) yields the famous *Bogomol'nyi bound* [26]:

$$M_M \geq |vg|. \quad (4.7)$$

This bound is saturated if and only if $V(\Phi) = 0$ (and, of course, $\vec{E} = 0$) and the Bogomol'nyi equation

$$\vec{B}^a = \vec{D}\Phi^a \quad (4.8)$$

is satisfied.¹⁴ It should be noticed that the *first-order* Bogomol'nyi equation implies the second-order equations of motion. The corresponding limit is known as the *Bogomol'nyi-Prasad-Sommerfield* (BPS) limit [25, 26]:

$$\vec{E}^a = 0, \quad D_0 \Phi^a = 0, \quad \vec{B}^a = \pm \vec{D} \Phi^a. \quad (4.9)$$

In quantum theory, where even the vanishing scalar potential may have radiative corrections, it is therefore important to protect flat directions of the potential, in order to achieve the Bogomol'nyi bound (see Part II).

After a substitution of the Ansatz (4.1) into the Bogomol'nyi equation (4.8), one finds

$$\xi \frac{dK}{d\xi} = -KH, \quad \xi \frac{dH}{d\xi} = H - (K^2 - 1), \quad (4.10)$$

whose solution is [25]

$$H(\xi) = \frac{\xi}{\tanh \xi} - 1, \quad K(\xi) = \frac{\xi}{\sinh \xi}. \quad (4.11)$$

When inserting this solution into eq. (4.3), one finds that the energy density is concentrated in the small region around the origin (i.e. in the core of a monopole). At distances greater than a Compton wavelength $(ve)^{-1} \sim M_W^{-1}$, where M_W is the mass of the W^\pm gauge particles resulting from the spontaneous symmetry breaking, the function K exponentially vanishes. In physical terms, it means that there is a cloud of W^\pm fields around the monopole while, well outside the monopole core, the magnetic field falls like r^{-2} , thus leaving that field configuration to be indistinguishable from the Dirac monopole. The Higgs fields also exponentially decay at spacial infinity, but they also have a long-range piece falling as r^{-1} :

$$\Phi^a \xrightarrow{r \rightarrow \infty} v e_r^a - \frac{e_r^a}{er}. \quad (4.12)$$

The presence of the last term follows from the Nambu-Goldstone theorem which predicts a massless 'dilaton' field D associated to the spontaneous breakdown of scale invariance. The field D can be introduced as [6]

$$\Phi^a = v e_r^a e^D, \quad (4.13)$$

while its dimensionless D -charge is given by

$$Q_D \equiv v \int_{S_\infty^2} \vec{\nabla} D \cdot d\vec{S} = \frac{4\pi}{e} = g = \frac{M_M}{v}. \quad (4.14)$$

¹⁴Writing $\Phi \equiv A_4$, the Bogomol'nyi equation in R^3 can be rewritten as the *self-dual* Yang-Mills equation in Euclidean space R^4 : $F_{ab} = {}^*F_{ab}$, where $a, b = 1, 2, 3, 4$, and all the fields are supposed to be independent upon x_4 .

One can conveniently describe the field of a point monopole by a *dual* potential $\tilde{A}^\mu = (\tilde{A}^0, \vec{\tilde{A}})$ defined by $*F = d\tilde{A}$. A coupling of another point monopole (of mass M) to this field is described by the action which is ‘dual’ to the standard action for an electrically charged point particle in an electro-magnetic field,

$$I_M = \int dt \left(-M\sqrt{1 - \vec{v}^2} - g\tilde{A}^0 + g\vec{v} \cdot \vec{\tilde{A}} \right), \quad (4.15)$$

where \vec{v} is a velocity of the test monopole. A sum of the standard electromagnetic action, $I_{\text{e.m.}} = -\frac{1}{4} \int d^4x *F*F$, and eq. (4.15) defines the total action which gives rise to a Coulomb magnetic field for a monopole at rest in the origin, as well as to the standard Coulomb repulsion between like sign monopoles, as it should have been expected from the dual picture. This picture should however be corrected since the theory also includes the massless ‘dilaton’ field D , as we already know, with the free action $I_D = \frac{v^2}{2} \int d^4x \partial_\mu D \partial^\mu D$. Accordingly, the full action must also include the coupling to the ‘dilaton’ field, which is dictated by the fact that a shift of the ‘dilaton’ field is equivalent to a shift in the mass of the monopole (scale invariance is spontaneously broken !). The action (4.15) should therefore be modified as [27]

$$I_{M,D} = \int dt \left(-[M + vD]\sqrt{1 - \vec{v}^2} - g\tilde{A}^0 + g\vec{v} \cdot \vec{\tilde{A}} \right). \quad (4.16)$$

The ultimate force between two stationary monopoles to be computed from that action turns out to be zero, which is consistent with the existence of multi-monopole static configurations [27].¹⁵ The space of solutions to the Bogomol’nyi equation (4.8) is called the *moduli space*, and it has dimension $4m$ [28]. Amongst the $4m$ *moduli* parameterizing the moduli space, $3m$ are just the space coordinates of the monopole locations, whereas the rest (m) corresponds to the monopole excitations of the electrically charged W^\pm fields in the core of the monopole.

In quantum theory, the classical BPS solution corresponds to a new particle – a *BPS state* – which is not present in the perturbative spectrum of the quantized Georgi-Glashow model, and whose mass is proportional to the *inverse* of the gauge coupling constant e , according to eq. (4.3). The last remark also explains why this BPS state cannot be seen in the weak coupling limit – simply because the mass of this state becomes very large when $e \rightarrow 0$.

The electro-magnetic duality (2.21) implies a generalization of the Bogomol’nyi bound (4.7) to the form

$$M_D \geq v\sqrt{q^2 + g^2}, \quad (4.17)$$

¹⁵There is, of course, a non-zero interaction between a monopole and an anti-monopole.

which applies to *dyons* having both a magnetic charge g and an electric charge q . In order to verify eq. (4.17), one has to construct a dyon solution. It was found by allowing a nonvanishing electric potential of the type [29]

$$A_0^a = \frac{e_r^a}{er} J(r) , \quad (4.18)$$

in the Ansatz (4.1). The equations of motion are therefore modified, namely,

$$\begin{aligned} \xi^2 \frac{d^2 K}{d\xi^2} &= K (K^2 + H^2 - J^2 - 1) , \\ \xi^2 \frac{d^2 H}{d\xi^2} &= 2K^2 H + \frac{\lambda}{e^2} H (H^2 - \xi^2) , \\ \xi^2 \frac{d^2 J}{d\xi^2} &= 2K^2 J , \end{aligned} \quad (4.19)$$

whose solution *in the BPS limit* $\lambda \rightarrow 0$ reads [29]

$$\begin{aligned} H(\xi) &= \cosh \gamma \left(\frac{\xi}{\tanh \xi} - 1 \right) , \\ K(\xi) &= \frac{\xi}{\sinh \xi} , \\ J(\xi) &= \sinh \gamma \left(\frac{\xi}{\tanh \xi} - 1 \right) , \end{aligned} \quad (4.20)$$

where γ is an arbitrary constant. The charges and the mass of this classical object are given by (*cf.* eq. (4.6))

$$q \equiv \frac{1}{v} \int d^3x D_i \Phi^a E_i^a = \frac{4\pi}{e} \sinh \gamma , \quad g \equiv \frac{1}{v} \int d^3x D_i \Phi^a B_i^a = \frac{4\pi}{e} , \quad (4.21)$$

and

$$M = \frac{4\pi}{e} v \cosh^2 \gamma . \quad (4.22)$$

It is now easy to verify the bound in eq. (4.17). It is remarkable that the BPS mass formula

$$M_{\text{BPS}} = v \sqrt{q^2 + g^2} , \quad (4.23)$$

does not distinguish between the ‘fundamental’ quantum particles and the monopoles, being applicable to all of them, like the meson-soliton democracy in the SG model. Semiclassical quantization of the dyon solution leads to the electric charge quantization [30] (see also the next sect. 5),

$$q = en_e , \quad \text{where } n_e \in \mathbf{Z} . \quad (4.24)$$

Thus, we just learned in this section that

- the BPS limit implies the existence of new BPS states in the quantum theory, which are absent in the perturbative spectrum,
- the BPS mass formula (4.23) is universal, and it is unvariant under the electromagnetic duality,
- the Coulomb repulsion between like sign static monopoles is exactly cancelled by the dilaton attraction.

5 Witten effect and S duality

The previous discussion of the electro-magnetic duality, although being supported by the BPS mass formula and the moduli space structure, still leaves many questions to be unanswered within the framework of the Georgi-Glashow model. For instance, the quantum Georgi-Glashow model cannot be duality invariant, since there are quantum corrections to all masses, which are not under control in that model. Also, since the W -bosons have spin, the magnetic monopoles should also have spin, whose origin in the Georgi-Glashow model is unclear. One still needs an underlying theory for describing dyons. The necessary additional input is provided by the extended supersymmetry¹⁶ and the so-called θ -term (or *vacuum angle*), which can be added to the Yang-Mills Lagrangian without spoiling its renormalizability,

$$\mathcal{L}_\theta = -\frac{\theta e^2}{32\pi^2} F_{\mu\nu}^a * F^{a\mu\nu} . \quad (5.1)$$

Being a total derivative, it does not affect the classical equations of motion, it violates P and CP , but not C , which makes it as a good candidate for generalizing the long-range behaviour of the theory while maintaining duality.

As was first noticed by Witten [31], the allowed values of electric charge in the monopole sector of the theory become shifted by the θ -term. For instance, an electromagnetic field in the presence of a Dirac monopole takes the form

$$\vec{E} = \vec{\nabla} A_0 , \quad \vec{B} = \vec{\nabla} \times \vec{A} + \frac{g}{4\pi} \frac{\vec{e}_r}{r^2} . \quad (5.2)$$

Its substitution into eq. (5.1) yields

$$\begin{aligned} L_\theta &= \int d^3r \mathcal{L}_\theta = \frac{\theta e^2}{8\pi^2} \int d^3r \vec{\nabla} A_0 \cdot \left(\vec{\nabla} \times \vec{A} + \frac{g}{4\pi} \frac{\vec{e}_r}{r^2} \right) \\ &= -\frac{\theta e^2 g}{32\pi^3} \int d^3r A_0 \vec{\nabla} \cdot \frac{\vec{e}_r}{r^2} = -\frac{\theta e^2 g}{8\pi^2} \int d^3r A_0 \delta^3(\vec{r}) , \end{aligned} \quad (5.3)$$

¹⁶The supersymmetry is discussed in the Part II.

which is just the coupling of the scalar potential A_0 to an *electric* charge of magnitude $-\theta e^2 g / (8\pi^2)$ located at the origin. In other words, the magnetic monopole has acquired an electric charge !

A more fundamental derivation of the same fact is based on the full spontaneously broken gauge theory with θ -term, whose total Lagrangian \mathcal{L}_{tot} is given by the sum of eqs. (3.1) and (5.1) in the BPS limit [31]. By the full theory here one means the non-local theory where magnetically neutral particles occur as quantum excitations of the fields present in the action, whereas magnetically charged particles (the BPS states) occur as solitons. Consider now a gauge transformation from the unbroken $U(1)$ subgroup (i.e. about the axis $e_{\Phi}^a \equiv \Phi^a / |\Phi^a|$) with the gauge parameter approaching a constant at spacial infinity. In the infinitesimal form, it is given by

$$\delta A_{\mu}^a = \frac{1}{ve} (D_{\mu} \Phi)^a , \quad (5.4)$$

where Φ^a is the background monopole Higgs field. Let \mathcal{N} be the generator of that gauge transformation. Its explicit form can be easily computed via the Noether method,

$$\mathcal{N} = \frac{\partial \mathcal{L}_{\text{tot}}}{\partial (\partial_0 A_{\mu}^a)} \delta A_{\mu}^a = \frac{q}{e} + \frac{\theta e g}{8\pi^2} , \quad (5.5)$$

where we have used δA_{μ}^a as of eq. (5.4), as well as the definitions of the total magnetic and electric charges in eq. (4.21). Since the rotation by 2π about the axis e_{Φ}^a should be trivial on physical states, we must have

$$\exp(2\pi i \mathcal{N}) = 1 , \quad \text{or, equivalently, } \mathcal{N} = n_e \in \mathbf{Z} . \quad (5.6)$$

Together with the Dirac quantization condition, $eg = 4\pi n_m$ where $n_m \in \mathbf{Z}$, eq. (5.5) now implies

$$q = en_e - \frac{e\theta}{2\pi} n_m , \quad (5.7)$$

which is a generalization of eq. (4.24). The *Witten effect* described by eq. (5.7) provides the physical meaning to the shift $\theta \rightarrow \theta + 2\pi$ which changes the induced electric charge of the BPS monopole.

It was the original *Montonen-Olive conjecture* [33] that the Georgi-Glashow model, i.e. the $SO(3)$ Yang-Mills-Higgs theory in the BPS limit (at $\theta = 0$) has an exact duality symmetry under the exchange of the fields, $\vec{E} \rightarrow \vec{B}$ and $\vec{B} \rightarrow -\vec{E}$, and the exchange of the coupling constants,

$$e \rightarrow g = \pm \frac{4\pi\hbar}{e} . \quad (5.8)$$

The dual or ‘magnetic’ formulation of the theory will also be a spontaneously broken gauge theory with essentially the same Lagrangian, where the W^{\pm} bosons would

appear as solitons, while the BPS monopoles would be ‘fundamental’. It is clear that eq. (5.8) represents a strong-weak coupling transformation, like that in eq. (1.10) for the SG–T quantum equivalence. Unfortunately, the corresponding ‘vertex operator construction’ connecting the two dual gauge theory formulations is not known in four dimensions.

The Montonen-Olive idea becomes extended when the θ -term is also taken into account. First, both coupling constants e and θ can be united into one *complex* parameter τ ,

$$\tau \equiv \frac{\theta}{2\pi} + \frac{4\pi i}{e^2} . \quad (5.9)$$

Since the physics is periodic in θ with period 2π , we have a duality transformation

$$T : \quad \tau \rightarrow \tau + 1 , \quad (5.10)$$

whereas the Montonen-Olive duality transformation (5.8) in terms of τ takes the form ¹⁷

$$S : \quad \tau \rightarrow -\frac{1}{\tau} . \quad (5.11)$$

It seems to be quite reasonable that the full duality symmetry is generated by the two transformations (5.10) and (5.11). They generate the group $SL(2, \mathbf{Z})$ of projective transformations

$$\tau \rightarrow \frac{a\tau + b}{c\tau + d} , \quad \text{where } a, b, c, d \in \mathbf{Z} , \quad \text{and } ad - bc = 1 . \quad (5.12)$$

Since $e^2 > 0$, the parameter τ naturally lives on the upper half plane, $\text{Im } \tau \geq 0$. Because of eqs. (5.7) and (5.9), the transformation (5.10) shifts the electric charge by -1 (for $n_m = 1$), while the transformation (5.11) exchanges electric and magnetic quantum numbers n_e and n_m . Putting all together, the action of the $SL(2, \mathbf{Z})$ on the quantum numbers reads

$$\begin{pmatrix} n_e \\ n_m \end{pmatrix} \rightarrow \begin{pmatrix} a & -b \\ c & -d \end{pmatrix} \begin{pmatrix} n_e \\ n_m \end{pmatrix} . \quad (5.13)$$

When being rewritten in terms of τ , the BPS bound $M^2 \geq v^2(Q_e^2 + Q_m^2)$, where $Q_e \equiv en_e - \frac{\theta e}{2\pi} n_m$ and $Q_m \equiv \frac{4\pi}{e} n_m$, takes the form

$$M^2 \geq 4\pi v^2(n_e, n_m) \frac{1}{\text{Im } \tau} \begin{pmatrix} 1 & -\text{Re } \tau \\ -\text{Re } \tau & |\tau|^2 \end{pmatrix} \begin{pmatrix} n_e \\ n_m \end{pmatrix} , \quad (5.14)$$

¹⁷Note that it is the τ , not e , which is inverted.

which is $SL(2, \mathbf{Z})$ invariant ! Most of the key equations above can be conveniently represented in terms of new variables

$$a \equiv ve \ , \quad \text{and} \quad a_D \equiv \tau a \ . \quad (5.15)$$

In particular, eqs. (5.10) and (5.11) now read

$$\begin{aligned} T &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} : \begin{pmatrix} a_D \\ a \end{pmatrix} \rightarrow \begin{pmatrix} a + a_D \\ a \end{pmatrix} , \\ S &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} : \begin{pmatrix} a_D \\ a \end{pmatrix} \rightarrow \begin{pmatrix} -a \\ a_D \end{pmatrix} , \end{aligned} \quad (5.16)$$

while the BPS mass spectrum is given by

$$M_{\text{BPS}} = |an_e + a_D n_m| \ . \quad (5.17)$$

Since the mass formula should be duality invariant, the charge vector $q_m = (n_m, n_e)$ also gets transformed under $M \in SL(2, \mathbf{Z})$ to $q' = qM^{-1}$, where $q'_m = (n'_m, n'_e)$ are also integers. The *stable* BPS states are those for which n_m and n_e are relatively prime [32].

Therefore, we just learned that

- in the presence of the θ -term, the naive (Montonen-Olive) electro-magnetic duality becomes extended to the projective transformations $SL(2, \mathbf{Z})$.

The extension (5.12) of the Montonen-Olive duality is called *S-duality* [34]. The S-duality invariance is a very strong requirement in quantum field theory. In particular, it implies that the renormalization group trajectories (if the theory has a non-vanishing beta-function) must be confined in the fundamental region of $SL(2, \mathbf{Z})$ in the τ -plane. If the beta-function is vanishing, the S-duality implies that the partition function of the theory is modular invariant (i.e. it must be a modular form). The only known candidates for such a behaviour are given by the *finite* gauge theories with N=2 or N=4 extended supersymmetry. It is the extended supersymmetry that also explains from the fundamental point of view the Bogomol'nyi bound, and provides an exact quantum status to the BPS states. Therefore, in order to proceed further in our studies of duality, we need to learn more about the extended supersymmetry in the next Part II of the review.

PART II: INTRODUCING SUPERSYMMETRY

In this Part of the review, some aspects of supersymmetry, which are going to be relevant in the Part III, are discussed. The emphasis is made on the superspace approach to the supersymmetric gauge theories with $N = 1$ and $N = 2$ supersymmetry. The BPS bound is related to the central charges appearing in the $N = 2$ extended supersymmetry algebra. The field content and the classical component action of the $N = 4$ supersymmetric gauge theory, which is believed to be exactly self-dual under the S-duality, is given. As a pre-requisite to the Seiberg-Witten results to be discussed in Part III, the moduli space of the $N = 2$ *supersymmetric Yang-Mills* (SYM) theory, its renormalization and the *low-energy effective action* (LEEA) are introduced.

1 Supersymmetry algebras and their representations

The Lorentz group $SO(1, 3)$ has the covering group $SL(2, \mathbf{C})$. Accordingly, a four-component (complex and anticommuting) Dirac spinor Ψ_D is a reducible representation. One can introduce the irreducible two-component complex spinors ψ_α and $\bar{\chi}^{\dot{\alpha}} = (\chi_\alpha)^*$ instead,

$$\Psi_D = \begin{pmatrix} \psi_\alpha \\ \bar{\chi}^{\dot{\alpha}} \end{pmatrix}. \quad (1.1)$$

The two-component spinor indices are raised and lowered with the antisymmetric ε -tensors, which represent the charge conjugation matrix,

$$\varepsilon^{\alpha\beta} = \varepsilon^{\dot{\alpha}\dot{\beta}} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (1.2)$$

We use the notation $\psi\chi \equiv \psi^\alpha\chi_\alpha$ and $\bar{\psi}\bar{\chi} \equiv \bar{\psi}_{\dot{\alpha}}\bar{\chi}^{\dot{\alpha}}$, so that $(\psi\chi)^\dagger = \bar{\psi}\bar{\chi}$. A convenient representation for the 4×4 Dirac matrices is given by

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}, \quad (1.3)$$

where the 2×2 σ -matrices are defined by

$$(\sigma^\mu)_{\alpha\dot{\alpha}} = (\mathbf{1}, \vec{\sigma})_{\alpha\dot{\alpha}}, \quad (\bar{\sigma}^\mu)^{\dot{\alpha}\alpha} = \varepsilon^{\dot{\alpha}\dot{\beta}}\varepsilon^{\alpha\beta}(\sigma^\mu)_{\beta\dot{\beta}} = (\mathbf{1}, -\vec{\sigma})^{\dot{\alpha}\alpha}, \quad (1.4)$$

and $\vec{\sigma} = (\sigma^1, \sigma^2, \sigma^3)$ are Pauli matrices. The $\gamma_5 = -i\gamma^0\gamma^1\gamma^2\gamma^3$ is diagonal in this representation, while $\gamma_5 = 1$ for the upper two components of Ψ_D and $\gamma_5 = -1$ for

the lower two components. The two-component spinors can be identified with the Weyl (chiral) spinors. A Majorana spinor is defined by

$$\Psi_M = \begin{pmatrix} \psi_\alpha \\ \bar{\psi}^{\dot{\alpha}} \end{pmatrix} . \quad (1.5)$$

One finds

$$\sigma^\mu P_\mu = \begin{pmatrix} P^0 - P^3 & -P^1 + iP^2 \\ -P^1 - iP^2 & P^0 + P^3 \end{pmatrix} , \quad \text{so that} \quad \det \sigma^\mu P_\mu = P^\mu P_\mu . \quad (1.6)$$

The Lorentz transformations for two-component spinors are generated by

$$\begin{aligned} (\sigma^{\mu\nu})_\alpha^\beta &= \frac{1}{4} \left[\sigma_\alpha^\mu \bar{\sigma}^{\nu\dot{\beta}\beta} - (\mu \leftrightarrow \nu) \right] , \\ (\bar{\sigma}^{\mu\nu})^{\dot{\alpha}}_{\dot{\beta}} &= \frac{1}{4} \left[\bar{\sigma}^{\mu\dot{\alpha}\beta} \sigma_{\beta\dot{\beta}}^\nu - (\mu \leftrightarrow \nu) \right] . \end{aligned} \quad (1.7)$$

The N -extended supersymmetry (susy) algebra without central charges reads

$$\{Q_\alpha^I, \bar{Q}_{\dot{\alpha}J}\} = 2\sigma_{\alpha\dot{\alpha}}^\mu P_\mu \delta^I_J , \quad \{Q_\alpha^I, Q_\beta^J\} = \{\bar{Q}_{\dot{\alpha}I}, \bar{Q}_{\dot{\beta}J}\} = 0 , \quad (1.8)$$

where $I, J = 1, 2, \dots, N$.¹⁸ The *massive* susy irreducible representations (irreps) can be easily found by using Wigner's method of induced representations. Defining $P^\mu = (M, 0, 0, 0)$ and rescaling the charges, one can represent eq. (1.8) as *two* Clifford algebras, each having the form

$$\{a^I, (a^J)^\dagger\} = \delta^{IJ} . \quad (1.9)$$

Hence, without incorporating CPT invariance, the susy irrep over the spin- j 'vacuum' $|\Omega\rangle_j$ has dimension $(2j+1)2^{2N}$. There is always an equal number of bosons and fermions, all having the same mass. The maximal helicity gap amongst the states in the representation is N . For example, if $N = 1$ and $j = 0$, one arrives at a *chiral* $N = 1$ susy multiplet comprising a Majorana (or Weyl) spinor and a complex scalar (2 bosonic and 2 fermionic degrees of freedom). Similarly, if $N = 1$ and $j = 1/2$, one finds an $N = 1$ *vector* multiplet which can be represented in field theory by a vector field, a Dirac fermion and a real scalar (4 bosonic and 4 fermionic degrees of freedom). The minimal massive $N = 2$ multiplet has $2^4 = 16$ states, whereas in the $N = 4$ case the minimal number of states increases to $2^8 = 256$ while the spin 2 appears.

¹⁸It is assumed in what follows that N is either 1, 2 or 4.

As far as the *massless* susy irreps are concerned, the situation is different. When choosing a frame where $P^\mu = M(1, 0, 0, 1)$, one easily finds from eq. (1.8) that

$$\{Q_\alpha^I, \bar{Q}_{\dot{\alpha}J}\} = \begin{pmatrix} 0 & 0 \\ 0 & 4M \end{pmatrix} \delta^I_J . \quad (1.10)$$

Hence, one of the Clifford algebras can be trivially realized, which effectively reduces the number of creation and destruction operators by half. As a result, there are now only $(2j+1)2^N$ states in the multiplets. If the vacuum has helicity λ , the highest helicity state has $\lambda + \frac{1}{2}N$. Accordingly, it yields

$$\begin{aligned} N=1 : & \quad |\lambda\rangle, \quad |\lambda - 1/2\rangle ; \\ N=2 : & \quad |\lambda\rangle, \quad 2|\lambda - 1/2\rangle, \quad |\lambda - 1\rangle ; \\ N=4 : & \quad |\lambda\rangle, \quad 4|\lambda - 1/2\rangle, \quad 6|\lambda - 1\rangle, \quad 4|\lambda - 3/2\rangle, \quad |\lambda - 2\rangle . \end{aligned} \quad (1.11)$$

In a local field theory, which is *CPT* invariant, one has to append the states (1.11) with their *CPT* conjugates, unless they are already *CPT* invariant.

The extended susy algebra can be modified [35]:

$$\{Q_\alpha^I, \bar{Q}_{\dot{\alpha}J}\} = 2\sigma^\mu_{\alpha\dot{\alpha}} P_\mu \delta^I_J, \quad \{Q_\alpha^I, Q_\beta^J\} = \varepsilon_{\alpha\beta} Z^{IJ}, \quad \{\bar{Q}_{I\dot{\alpha}}, \bar{Q}_{J\dot{\beta}}\} = \varepsilon_{\dot{\alpha}\dot{\beta}} Z_{IJ}^*, \quad (1.12)$$

where the *central charges* $Z^{IJ} = -Z^{JI}$ have been introduced. In the $N=2$ case, they reduce to a single (complex) central charge, $Z^{IJ} = 2\varepsilon^{IJ}Z$, while Z can be fixed to be real by a chiral rotation. Defining

$$\begin{aligned} a_\alpha &= \frac{1}{\sqrt{2}} [Q_\alpha^1 + \varepsilon_{\alpha\beta} (Q_\beta^2)^\dagger], \\ b_\alpha &= \frac{1}{\sqrt{2}} [Q_\alpha^1 - \varepsilon_{\alpha\beta} (Q_\beta^2)^\dagger], \end{aligned} \quad (1.13)$$

one finds for massive representations that

$$\{a_\alpha, a_\beta^\dagger\} = 2(M + |Z|)\delta_{\alpha\beta}, \quad \{b_\alpha, b_\beta^\dagger\} = 2(M - |Z|)\delta_{\alpha\beta}, \quad (1.14)$$

while all the other anticommutators vanish. Eq. (1.14) leads to the bound

$$M \geq |Z|. \quad (1.15)$$

When this bound becomes saturated, $|Z| = M$, the massive representation becomes smaller, and one gets a *reduced* massive multiplet comprising the BPS states (sect. I.4). The reduction mechanism is quite similar to that for the massless susy representations without central charges, and it results in the same number of states at given N .

This fact is important for a consistency of the Higgs mechanism in supersymmetric gauge theories, which assumes an equal number of degrees of freedom before and after spontaneous gauge symmetry breaking. For example, a *reduced* or *short* massive $N = 2$ multiplet can have only two bosonic and two fermionic degrees of freedom. Similarly, there are only 16 states in the short massive representation of $N = 4$ supersymmetry.

The states that become massive by the Higgs mechanism must belong to *short* supermultiplets, as they were before the spontaneous symmetry breaking, since the Higgs mechanism cannot generate the extra massive states which appear in an unreduced ('long') massive supermultiplet.

One concludes that the status of BPS states in a short representation of extended supersymmetry is well defined in quantum theory, unless the supersymmetry is not broken, because

- the BPS states are protected by extended supersymmetry. Hence, their existence does not depend on the underlying dynamics of quantum theory. In particular, it remains to be true at strong coupling,
- The short massive supermultiplets of BPS states can be equally defined by requiring *a half* of the supersymmetry generators to vanish on them, in the presence of central charges.

2 $N = 1$ field theories and superspace

Since supersymmetry representations in field theory appear as multiplets comprising bosonic and fermionic fields, one needs their unified description. Such a description is provided by *superspace* [36, 37, 38, 39]. The basic idea of superspace is to extend spacetime by anticommuting coordinates that are spacetime spinors and whose number is just equal to the number of supersymmetry generators. The supersymmetry transformations can then be realized as certain translations in superspace, while a tensor in superspace automatically provides a supersymmetry representation. In the case of the unextended (*simple*) $N = 1$ supersymmetry, the $N = 1$ superspace coordinates are given by $z^M = (x^\mu, \theta^\alpha, \bar{\theta}^{\dot{\alpha}})$. The superspace realization of the supersymmetry generators (without central charges) reads

$$Q_\alpha = +\frac{\partial}{\partial\theta^\alpha} - i\sigma^\mu_{\alpha\dot{\alpha}}\bar{\theta}^{\dot{\alpha}}\partial_\mu, \quad \bar{Q}_{\dot{\alpha}} = -\frac{\partial}{\partial\bar{\theta}^{\dot{\alpha}}} + i\sigma^\mu_{\alpha\dot{\alpha}}\theta^\alpha\partial_\mu. \quad (2.1)$$

However, there is a problem since a general superfield provides a reducible representation of supersymmetry. Hence, one needs to develop a covariant calculus in superspace. The main tools are the spinorial covariant derivatives

$$D_\alpha = +\frac{\partial}{\partial\theta^\alpha} + i\sigma^\mu_{\alpha\dot{\alpha}}\bar{\theta}^{\dot{\alpha}}\partial_\mu, \quad \bar{D}_{\dot{\alpha}} = -\frac{\partial}{\partial\bar{\theta}^{\dot{\alpha}}} - i\sigma^\mu_{\alpha\dot{\alpha}}\theta^\alpha\partial_\mu, \quad (2.2)$$

which anticommute with the supersymmetry generators (2.1), and satisfy a similar algebra. The simplest chiral scalar multiplet is given by the *chiral* scalar superfield Φ satisfying the superspace constraint

$$\bar{D}_{\dot{\alpha}}\Phi = 0. \quad (2.3)$$

This constraint can be easily solved,¹⁹ $\Phi = \Phi(y, \theta) = \phi(y) + \sqrt{2}\theta\psi(y) + \theta^2 F(y)$, where $y^\mu = x^\mu + i\theta\sigma^\mu\bar{\theta}$, or, more explicitly,

$$\Phi = \phi(x) + \sqrt{2}\theta\psi(x) + \theta^2 F(x) + i\theta\sigma^\mu\bar{\theta}\partial_\mu\phi(x) - \frac{i}{\sqrt{2}}\theta^2(\partial_\mu\psi(x)\sigma^\mu\bar{\theta}) - \frac{1}{4}\theta^2\bar{\theta}^2\partial^2\phi(x). \quad (2.4)$$

The *antichiral* superfield Φ^\dagger satisfies $D_\alpha\Phi^\dagger = 0$, whereas a supersymmetry invariant action is simply given by a full superspace integral

$$\frac{1}{4}\int d^4x d^2\theta d^2\bar{\theta} \Phi^\dagger \Phi = \int d^4x (\partial_\mu\phi\partial^\mu\phi^\dagger - i\bar{\psi}\bar{\sigma}^\mu\partial_\mu\psi + F^\dagger F). \quad (2.5)$$

Obviously, any function of chiral superfields is again a chiral superfield, so that eq. (2.5) can be easily generalized to include interactions, whose most general form is

$$\mathcal{L} = \int d^4\theta K(\Phi, \bar{\Phi}) + \int d^2\theta W(\Phi) + \int d^2\bar{\theta} \bar{W}(\bar{\Phi}), \quad (2.6)$$

where a *Kähler* potential K and a holomorphic *superpotential* W have been introduced.

A general real scalar superfield V can be written down in the form [36]

$$\begin{aligned} V(x, \theta, \bar{\theta}) = & C + i\theta\chi - i\bar{\theta}\bar{\chi} + \frac{i}{2}\theta^2(M + iN) - \frac{i}{2}\bar{\theta}^2(M - iN) - \theta\sigma^\mu\bar{\theta}A_\mu \\ & + i\theta^2\bar{\theta}(\bar{\lambda} + \frac{i}{2}\bar{\sigma}^\mu\partial_\mu\chi) - i\bar{\theta}^2\theta(\lambda + \frac{i}{2}\sigma^\mu\partial_\mu\bar{\chi}) + \frac{1}{2}\theta^2\bar{\theta}^2(D - \frac{1}{2}\square C). \end{aligned} \quad (2.7)$$

This superfield is a reducible representation of supersymmetry, since it contains the smaller chiral and antichiral superfields, Λ and Λ^\dagger . They can be effectively removed by imposing the gauge symmetry

$$V \rightarrow V + i\Lambda - i\Lambda^\dagger. \quad (2.8)$$

¹⁹We use the conventional notation: $\theta^2 = \theta^\alpha\theta_\alpha$, $\theta\sigma^\mu\bar{\theta} = \theta^\alpha\sigma^\mu_{\alpha\dot{\alpha}}\bar{\theta}^{\dot{\alpha}}$, etc.

In the so-called *Wess-Zumino gauge*, one chooses $C = M = N = \chi = 0$, which reduces eq. (2.7) to

$$V = -\theta\sigma^\mu\bar{\theta}A_\mu + i\theta^2(\bar{\theta}\bar{\lambda}) - i\bar{\theta}^2(\theta\lambda) + \frac{1}{2}\theta^2\bar{\theta}^2D, \quad (2.9)$$

thus defining a *vector* multiplet comprising a massless gauge field A_μ , its superpartner (*gaugino*) λ_α and a real auxiliary field D . Note that $V^3 \equiv 0$ in the Wess-Zumino gauge.

The abelian gauge-invariant superfield strength is given by the chiral spinor superfield W_α ,

$$W_\alpha = -\frac{1}{4}\bar{D}^2D_\alpha V, \quad \bar{W}_{\dot{\alpha}} = -\frac{1}{4}D^2\bar{D}_{\dot{\alpha}}V, \quad (2.10)$$

satisfying the constraint $DW = \bar{D}\bar{W}$. In the Wess-Zumino gauge, it reads

$$W(y) = -i\lambda + \theta D - i\sigma^{\mu\nu}\theta(\partial_\mu A_\nu - \partial_\nu A_\mu) + \theta^2\sigma^\mu\partial_\mu\bar{\lambda}. \quad (2.11)$$

In the non-abelian theory, all the fields of the vector multiplet, as well as the corresponding superfields, are to be assigned in the adjoint, $A_\mu = A_\mu^a t^a$, $[t^a, t^b] = f^{abc}t^c$, $\text{tr}(t^a t^b) = 2\delta^{ab}$, etc. The non-abelian version of eqs. (2.8) and (2.10) actually follows from a solution to the constraints on the gauge-covariant and super-covariant spinorial derivatives, \mathcal{D}_α and $\bar{\mathcal{D}}_{\dot{\alpha}}$, defining the super-Yang-Mills theory in superspace [36]. Instead of going into detail, the form of the non-abelian solution can be anticipated from the abelian eqs. (2.8) and (2.10). For example, as regards the gauge transformations with a Lie-algebra valued chiral superfield parameter Λ , one finds that

$$e^{-2eV} \rightarrow e^{+i\Lambda^\dagger} e^{-2eV} e^{-i\Lambda}, \quad (2.12)$$

whereas, as far as the non-abelian superfield strength is concerned, it reads

$$\begin{aligned} W_\alpha &\equiv \frac{1}{8e}\bar{D}^2\left(e^{2eV}D_\alpha e^{-2eV}\right) = -\frac{1}{4}\bar{D}^2(D_\alpha V + e[V, D_\alpha V]), \\ &= -i\lambda + \theta D - i\sigma^{\mu\nu}\theta F_{\mu\nu} + \theta^2\sigma^\mu\nabla_\mu\bar{\lambda}, \end{aligned} \quad (2.13)$$

where the Wess-Zumino gauge has been used, while $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ie[A_\mu, A_\nu]$ and $\nabla_\mu\lambda = \partial_\mu\lambda - ie[A_\mu, \lambda]$ as usual. The gauge-invariant kinetic term for the matter chiral superfields in some (for example, the adjoint) representation is given by

$$\begin{aligned} I_{\text{matter}} &= \frac{1}{4}\int d^4x d^2\theta d^2\bar{\theta} \text{tr}(\Phi^\dagger e^{-2eV}\Phi) \\ &= \frac{1}{4}\int d^4x d^2\theta d^2\bar{\theta} \text{tr}\left(\Phi^\dagger\Phi - 2e\Phi^\dagger V\Phi + 2e^2\Phi^\dagger V^2\Phi\right), \\ &= \int d^4x \text{tr}\left(|\nabla_\mu\phi|^2 - i\bar{\psi}\bar{\sigma}^\mu\nabla_\mu\psi + F^\dagger F \right. \\ &\quad \left. - e\phi^\dagger[D, \phi] - ie\sqrt{2}\phi^\dagger\{\lambda, \psi\} + ie\sqrt{2}\bar{\psi}[\bar{\lambda}, \phi]\right), \end{aligned} \quad (2.14)$$

whereas the natural (complex) kinetic term for the gauge fields reads

$$-\frac{1}{4} \int d^4x d^2\theta \operatorname{tr} W^\alpha W_\alpha = \int d^4x \operatorname{tr} \left[-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{i}{4} F_{\mu\nu}^* F^{\mu\nu} - i\lambda \sigma^\mu \nabla_\mu \bar{\lambda} + \frac{1}{2} D^2 \right] . \quad (2.15)$$

In addition to the standard kinetic term for the Yang-Mills field, eq. (2.15) also contains the θ -term, as required by supersymmetry. We are therefore guided by supersymmetry to introduce the complex coupling constant τ as in eq. (I.5.9), and then define the following *real* action:

$$\begin{aligned} I_{\text{SYM}} &= \frac{1}{16\pi} \operatorname{Im} \left[\tau \int d^4x d^2\theta \operatorname{tr} W^\alpha W_\alpha \right] \\ &= \frac{1}{e^2} \int d^4x \operatorname{tr} \left[-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - i\lambda \sigma^\mu \nabla_\mu \bar{\lambda} + \frac{1}{2} D^2 \right] + \frac{\theta}{32\pi^2} \int d^4x F_{\mu\nu}^* F^{\mu\nu} . \end{aligned} \quad (2.16)$$

It can be shown that the non-abelian superfield strength W_α is (i) a covariantly chiral superfield, $\bar{\mathcal{D}}_\alpha W_\alpha = \mathcal{D}_\alpha \bar{W}^\alpha = 0$, and (ii) satisfies the constraint

$$\mathcal{D}^\alpha W_\alpha = \bar{\mathcal{D}}_\alpha \bar{W}^\alpha . \quad (2.17)$$

These two conditions actually *define* the $N = 1$ super-Yang-Mills theory in superspace, and determine the component content of the theory in the Wess-Zumino gauge, as given above.

3 $N=2$ super-Yang-Mills theory

The most natural framework for the $N = 2$ extended supersymmetry is provided by $N = 2$ superspace, whose coordinates $z^M = (x^\mu, \theta_i^\alpha, \bar{\theta}_i^{\dot{\alpha}})$ contain two sets of the anticommuting spinor variables ($i = 1, 2$) related to each other by internal symmetry rotations. The $N = 2$ super-Yang-Mills (SYM) theory in $N = 2$ superspace can be defined by imposing appropriate constraints on the gauge-covariant and super-covariant spinorial derivatives \mathcal{D}_α^i and $\bar{\mathcal{D}}_{\dot{\alpha}}^i$ [40]. The constraints essentially amount to the existence of the $N = 2$ SYM field strength – a covariantly chiral scalar $N = 2$ superfield Ψ – satisfying the reality condition

$$\mathcal{D}_i^\alpha \mathcal{D}_{\alpha j} \Psi = \bar{\mathcal{D}}_{\dot{\alpha} i} \bar{\mathcal{D}}_{\dot{\alpha} j} \bar{\Psi} , \quad (3.1)$$

which is analogous to eq. (2.17). However, unlike in the $N = 1$ case, an $N = 2$ supersymmetric solution to the $N = 2$ non-abelian superspace constraints is not

known in an analytic form.²⁰ Therefore, instead of discussing the $N = 2$ constraints and their solution in $N = 2$ superspace, we are going to make a ‘short cut’, and first construct the $N = 2$ SYM theory in terms of $N = 1$ superfields.

Since the on-shell field content of an $N = 2$ vector multiplet is given by a sum of an $N = 1$ vector multiplet and a chiral $N = 1$ scalar multiplet, in the Wess-Zumino gauge, where the super-gauge degrees of freedom are eliminated, we should expect the gauge-covariant $N = 2$ SYM field strength Ψ be expressible in terms of the $N = 1$ gauge-covariant superfields Φ and W_α , all in the adjoint representation of the gauge group. Expanding the $N = 2$ covariantly chiral superfield Ψ in terms of a ‘half’ of proper chiral anticommuting coordinates,

$$\Psi = \Phi + \sqrt{2}\Theta^\alpha W_\alpha + \Theta^2 G , \quad (3.2)$$

we can represent Ψ in terms of three gauge-covariant $N = 1$ chiral superfields, Φ , W_α and G . Using dimensional reasons, we can now identify the $N = 1$ superfields Φ and W_α with the superfields appearing in eqs. (2.4) and (2.13), respectively. The remaining $N = 1$ superfield G is expected to be a (complicated) gauge-covariant chiral function of Φ and V , whose explicit form we do not need [5].

As far as the action of the $N = 2$ SYM theory is concerned, it should be given by a sum of eqs. (2.14) and (2.16) with proper relative normalization.²¹ Hence, the $N = 2$ SYM action in $N = 1$ superspace reads as follows:

$$\begin{aligned} I_{N=2 \text{ SYM}} &= \int d^4x \left[\text{Im} \left(\frac{\tau}{16\pi} \int d^2\theta \text{tr} W^\alpha W_\alpha \right) + \frac{1}{4e^2} \int d^2\theta d^2\bar{\theta} \text{tr} \Phi^\dagger e^{-2eV} \Phi \right] , \\ &= \text{Im} \text{tr} \int d^4x \frac{\tau}{16\pi} \left[\int d^2\theta W^\alpha W_\alpha + \int d^2\theta d^2\bar{\theta} \Phi^\dagger e^{-2eV} \Phi \right] . \end{aligned} \quad (3.3)$$

When dealing with an $N = 2$ theory in $N = 1$ superspace, one does not take care of the underlying off-shell $N = 2$ supersymmetry structure of the $N = 2$ theory, while the on-shell physics is of course the same. It is also possible to write down the $N = 2$ SYM action in $N = 2$ superspace. The $N = 2$ action should have the form of a chiral integral (on dimensional reasons), and the only gauge-invariant candidate is given by the trace of the $N = 2$ SYM superfield strength Ψ squared. The correct answer reads

$$I_{N=2 \text{ SYM}} = \text{Im} \left(\frac{\tau}{16\pi} \int d^4x d^4\theta \frac{1}{2} \text{tr} \Psi^2 \right) . \quad (3.4)$$

²⁰The $N = 2$ analogue of the V -superfield is given by an unconstrained $N = 2$ tensor superfield V_{ij} of dimension -2 . An analytic relation between Ψ and V_{ij} is not known in the non-abelian case.

²¹The relative normalization is easily fixed by requiring all fermionic kinetic terms to have the same coefficients.

The $N = 2$ SYM action in components can be easily recovered from eqs. (2.14), (2.16) and (3.3). In particular, the structure of auxiliary fields is governed by the action

$$I_{\text{aux}} = \frac{1}{e^2} \int d^4x \operatorname{tr} \left[\frac{1}{2} D^2 - e \phi^\dagger [D, \phi] + F^\dagger F \right] . \quad (3.5)$$

Eliminating the auxiliary fields D and F via their algebraic equations of motion yields

$$I_{\text{aux}} = -\frac{1}{2} \int d^4x \operatorname{tr} \left([\phi^\dagger, \phi] \right)^2 . \quad (3.6)$$

The potential $V(\phi) \equiv \frac{1}{2} \operatorname{tr}([\phi^\dagger, \phi])^2$ is therefore non-negative, but has *flat* directions. The non-trivial solutions to the equation $V(\langle \phi \rangle) = 0$ follow from

$$[\langle \phi^\dagger \rangle, \langle \phi \rangle] = 0 , \quad \langle \phi \rangle \neq 0 , \quad (3.7)$$

or, equivalently,

$$[\langle S \rangle, \langle P \rangle] = 0 , \quad (3.8)$$

where the scalar S and the pseudo-scalar P have been introduced, $\phi \equiv \frac{1}{\sqrt{2}}(S + iP)$. The parity-conserving solution to eq. (3.8) in the $SU(2)$ case is

$$\langle S^a \rangle = v \delta^{a3} , \quad \langle P^a \rangle = 0 , \quad (3.9)$$

where the value of the real parameter v is arbitrary. The set of all solutions to eq. (3.7) modulo gauge transformations is the classical moduli space of the theory, which is parametrized by the gauge-invariant parameter $\operatorname{tr} \langle \phi^2 \rangle = \frac{1}{2} v^2$ (see sect. 5 for more).

The $N = 2$ SYM Lagrangian in components can be written down in the form

$$\begin{aligned} \mathcal{L}_{N=2 \text{ SYM}} = & \frac{1}{4\pi} \operatorname{Im} \left\{ \left(\frac{\theta}{2\pi} + i \frac{4\pi}{e^2} \right) \operatorname{tr} \left[-\frac{1}{4} \left(F_{\mu\nu} F^{\mu\nu} - \frac{1}{2} \varepsilon^{\mu\nu\rho\lambda} F_{\mu\nu} F_{\rho\lambda} \right) \right. \right. \\ & \left. \left. + (D_\mu \phi)^\dagger (D_\mu \phi) - \frac{1}{2} ([\phi, \phi^\dagger])^2 + \dots \right] \right\} , \end{aligned} \quad (3.10)$$

where the scalar and spinor component fields have been rescaled, and the dots stand for fermionic terms. In the $SU(2)$ case, eq. (3.10) has the structure which is very similar to that of the Georgi-Glashow model, except of the potential. The $N = 2$ SYM action is classically scale (and conformally) invariant, but this invariance is spontaneously broken, if $\langle \phi \rangle \neq 0$. Unbroken supersymmetry requires the vanishing vacuum expectation values for all the auxiliary fields and, hence, implies $V(\langle \phi \rangle) = 0$.

With $SU(2)$ as the SYM gauge group, eq. (3.9) at $v \neq 0$ spontaneously breaks it down to $U(1)$. The BPS monopole solution (Part I) can be embedded into the $N = 2$ SYM theory, whose fields S^a replace Φ^a overthere and satisfy the Bogomol'nyi

bound $B_i^a = D_i S^a$. Unlike the Georgi-Glashow model, the BPS limit in the $N = 2$ SYM theory can be reached without sending the potential coupling constant to zero.

One can check whether a charge-one monopole solution has some supersymmetry. Since the fermionic fields have to vanish initially, their supersymmetry variations have to vanish too. The $N = 2$ supersymmetry variation of gaugino's in the BPS limit is governed by the operator

$$(\sigma^{\mu\nu} F_{\mu\nu} - \gamma^\mu \nabla_\mu S) \varepsilon = \gamma^i B_i (1 - \gamma_5) \varepsilon = 0 , \quad (3.11)$$

which implies that a chiral *half* of the supersymmetry remains unbroken.²²

As was shown in sect. 1, the $N = 2$ extended supersymmetry algebra can be modified by the inclusion of central charges. In an $N = 2$ supersymmetric field theory, the supersymmetry charges are expressed as space integrals of supersymmetry currents given by certain polynomials in fields and their derivatives. In the presence of monopoles carrying magnetic charges, the central terms in the $N = 2$ supersymmetry algebra of the $N = 2$ SYM theory can therefore be explicitly calculated. It was done by Olive and Witten [41] who found that

$$\begin{aligned} \text{Re } Z &= \int d^3x \partial_i [S^a E_i^a + P^a B_i^a] = v Q_e , \\ \text{Im } Z &= \int d^3x \partial_i [P^a E_i^a + S^a B_i^a] = v Q_m , \end{aligned} \quad (3.12)$$

where eq. (3.9) has been used, as well as the definitions of the total electric and magnetic charges (Part I). Hence, one gets

$$Z = v(Q_e + iQ_m) \quad \text{or, equivalently,} \quad |Z|^2 = v^2(Q_e^2 + Q_m^2) , \quad (3.13)$$

as well as the Bogomol'nyi bound $M \geq |Z| = v\sqrt{Q_e^2 + Q_m^2}$ as the direct consequences of extended supersymmetry ! Inverting the argument, the Bogomol'nyi equation follows by demanding the monopole solution be annihilated by half of the supersymmetries (i.e. form a *short* representation of $N = 2$ supersymmetry). Therefore, if $N = 2$ supersymmetry is not dynamically broken in quantum theory, the Bogomol'nyi bound is not going to be modified by quantum corrections, and the BPS states with magnetic charges will occur in the full quantum $N = 2$ SYM theory as well. If the 'fundamental' particles get their masses via the Higgs mechanism which does not change the number of physical degrees of freedom, they also fall into reduced (short) representations of $N = 2$ supersymmetry, and they can therefore also be considered as the BPS states.

²²The non-vanishing (of opposite chirality) supersymmetry variations of gaugino's are Dirac's zero-modes in the monopole background.

Assuming that the $N = 2$ supersymmetry of the classical $N = 2$ SYM theory is maintained in the full quantum theory,²³ it is possible to predict the form of its *low-energy effective action*,²⁴

$$I_{\mathcal{F}} = \frac{1}{16\pi} \text{Im} \int d^4x d^4\theta \mathcal{F}(\Psi) , \quad (3.14)$$

where \mathcal{F} is a holomorphic function, called the $N = 2$ *prepotential*. The classical part of the $N = 2$ prepotential is dictated by eq. (3.10):

$$\mathcal{F}_{\text{class}}(\Psi) = \frac{1}{2} \text{tr} \tau_{\text{cl}} \Psi^2 . \quad (3.15)$$

where τ_{cl} is given by eq. (I.5.9). The quadratic dependence in eq. (3.15) is crucial for renormalisability. In $N = 1$ superspace, the $N = 2$ SYM low-energy effective action (3.14) reads as follows [43]:

$$I_{\mathcal{F}} = \frac{1}{16\pi} \text{Im} \int d^4x \left[\int d^2\theta \mathcal{F}_{ab}(\Phi) W^{a\alpha} W_{\alpha}^b + \int d^2\theta d^2\bar{\theta} \left(\Phi^{\dagger} e^{-2eV} \right)^a \mathcal{F}_a(\Phi) \right] , \quad (3.16)$$

where we have used the notation

$$\mathcal{F}_a(\Phi) \equiv \partial \mathcal{F}(\Phi) / \partial \Phi^a , \quad \mathcal{F}_{ab}(\Phi) \equiv \partial^2 \mathcal{F}(\Phi) / \partial \Phi^a \partial \Phi^b . \quad (3.17)$$

One concludes that

- the BPS condition which was initially found at the classical level (Part I) is maintained in the full quantum theory as well, because it is a consequence of the extended supersymmetry,
- the mass formula for the BPS states (see e.g., the right-hand-side of eq. (I.5.14)) is *exact*, i.e. it holds in the full quantum theory, and it is valid for all particles in the semiclassical spectrum,
- the low-energy effective action of the $N = 2$ SYM theory is governed by a holomorphic prepotential \mathcal{F} .

The holomorphic function \mathcal{F} is expected to receive both perturbative and non-perturbative contributions after quantization. The tools to calculate the $N = 2$ prepotential exactly, by using a non-trivial interplay between holomorphicity, extended supersymmetry and duality, will be provided in Part III.

²³The Witten index does not vanish for the $N = 2$ SYM theory, which means that the $N = 2$ supersymmetry cannot be dynamically broken in that theory [42].

²⁴The low-energy part of the full (non-local) effective action represents the component kinetic terms with no more than two derivatives, and no more than four-fermion couplings.

4 N=4 super-Yang-Mills theory

Though the $N = 4$ super-Yang-Mills theory can be formulated on-shell in the conventional N=4 superspace, it is very difficult to construct its off-shell $N = 4$ supersymmetric formulation, if any. Therefore, we are going to confine ourselves to its component formulation. The easiest way to construct the four-dimensional $N = 4$ SYM theory is provided by dimensional reduction of the *ten-dimensional* supersymmetric gauge theory down to four dimensions [44].

The main point here is related to the dimension of a spinor representation in various space-time dimensions. The number of on-shell bosonic degrees of freedom in the case of a real vector gauge field A_M in D dimensions is $D - 2$, while the (real) number of on-shell fermionic degrees of freedom in the case of a Dirac spinor λ is $2^{[D/2]}$. Either the Weyl or the Majorana condition on λ reduces the last number by a factor of $1/2$. Therefore, the maximal dimension where the numbers of bosonic and fermionic degrees of freedom match for a minimal vector supermultiplet comprising (A_M, λ) is $D = 10$ *provided* that λ is Majorana and Weyl simultaneously, which is allowed in ten dimensions.²⁵

The action of the supersymmetric Yang-Mills theory in ten dimensions reads

$$I_{10} = \int d^{10}x \operatorname{tr} \left[-\frac{1}{4} F_{MN} F^{MN} - \frac{1}{2} \bar{\lambda} \Gamma^M (D_M \lambda) \right] , \quad (4.1)$$

where both fields A_M^a and λ^a are in the adjoint of the gauge group, and

$$(1 - \Gamma_{11})\lambda = 0 , \quad \bar{\lambda} = \lambda^T C_{10} . \quad (4.2)$$

We use the standard notation:

$$F_{MN}^a = \partial_M A_N^a - \partial_N A_M^a - e f^{abc} A_M^b A_N^c , \quad (D_M \lambda)^a = \partial_M \lambda^a - e f^{abc} A_M^b \lambda^c , \quad (4.3)$$

as usual. In eq. (4.2) one has $\Gamma_{11} = \Gamma_0 \Gamma_1 \Gamma_2 \cdots \Gamma_9$, while C_{10} is the charge conjugation matrix in ten dimensions, $C_{10} \Gamma_M C_{10}^{-1} = -\Gamma_M^T$. The early lower-case Latin letters are still used for the gauge group indices, while the capital Latin letters, $M, N, \dots = 0, 1, \dots, 9$, are used to denote the Lorentz indices in ten dimensions. It is straightforward to verify that the action (4.1) is invariant under the supersymmetry transformations

$$\delta A_M^a = \bar{\varepsilon} \Gamma_M \lambda^a , \quad \delta \lambda^a = -\sigma^{MN} F_{MN}^a \varepsilon , \quad (4.4)$$

²⁵Similarly, the $N = 2$ SYM theory can be obtained by dimensional reduction from the supersymmetric gauge theory in $D = 6$ provided that the superpartner of the Yang-Mills field is a Weyl spinor in the adjoint representation of the gauge group [44].

where the infinitesimal supersymmetry parameter ε is also a Majorana-Weyl spinor, and $\sigma^{MN} = \frac{1}{4}[\Gamma^M, \Gamma^N]$.

The dimensional reduction essentially amounts to requiring all the fields be only dependent on the four-dimensional space-time coordinates x^μ , while $x^M = (x^m, x^i)$ and $\mu = 0, 1, 2, 3$. From the group-theoretical viewpoint, it reduces the Lorentz group $SO(1, 9)$ to $SO(1, 3) \otimes SO(6)$. As a result, the fermionic field λ decomposes off-shell as $16 = (2_+, 4_+) + (2_-, 4_-)$, where the subscripts denote the space-time chirality. The ten-dimensional Dirac matrices can also be represented in terms of the four-dimensional Dirac matrices and some internal 4×4 matrices. Similarly, the gauge fields are decomposed off-shell as $10 = (4, 1) + (1, 6)$, which leads to a gauge field, three scalars and three pseudo-scalars, all in the adjoint, in four dimensions. Because of the isomorphism $Spin(6) \equiv SU(4)$, the resulting four-dimensional Lagrangian can be written in various forms. For instance, the six scalar fields can be united into an antisymmetric complex matrix ϕ_{ij} subject to the $SU(4)$ self-duality condition

$$\phi_{ij}^\dagger = \phi^{ij} = \frac{1}{2}\varepsilon^{ijkl}\phi_{kl} , \quad (4.5)$$

where $i, j, \dots = 1, 2, 3, 4$. As a result, the Lagrangian of the $N = 4$ SYM theory, which follows from eq. (4.1) after the dimensional reduction, is given by

$$\begin{aligned} \mathcal{L}_{N=4 \text{ SYM}} = & \text{tr} \left(-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\lambda_i\sigma^\mu D_\mu\bar{\lambda}^i + \frac{1}{2}D_\mu\phi_{ij}D^\mu\phi^{ij} \right. \\ & \left. + i\lambda_i[\lambda_j, \phi^{ij}] + i\bar{\lambda}^i[\bar{\lambda}^j, \phi_{ij}] + \frac{1}{4}[\phi_{ij}, \phi_{kl}][\phi^{ij}, \phi^{kl}] \right) . \end{aligned} \quad (4.6)$$

The $N = 4$ SYM theory also has monopole and dyon solutions, similar to the $N = 2$ SYM theory [45]. In the $N = 4$ theory, it is actually possible to have monopoles carrying spin 1, which overcomes one of the obstacles mentioned in Part I. Indeed, since there is a *unique* $N = 4$ multiplet with the highest spin 1, the monopole $N = 4$ supermultiplet must be isomorphic to the $N = 4$ gauge supermultiplet, have 16 states, and one state of spin 1, in particular.²⁶ Moreover, the $N = 4$ SYM theory is known to be *UV-finite* [46, 47, 48], i.e. it has vanishing beta-function and it is exactly scale invariant. Altogether, it selects the $N = 4$ SYM theory as a good candidate which may support the *exact* Montonen-Olive duality. In the $N = 2$ SYM theory, the S-duality can only be *effective*, not exact, being a subgroup of $SL(2, \mathbf{Z})$ (see Part III for details).

²⁶In the $N = 2$ SYM theory, the monopole solution belongs to a hypermultiplet [45], which does not contain a spin-1 state.

5 Moduli space of the $N = 2$ SYM theory

The $N = 2$ SYM scalar potential has flat directions to be determined as solutions to eq. (3.7). All the vacuum field configurations define the vacuum ‘manifold’ (Part I) which is parametrized by the vacuum expectation values of the scalar (Higgs) field. Since the vacua related by a gauge transformation describe the same physics, we are interested in the gauge-inequivalent vacua forming the *moduli space* \mathcal{M} and corresponding to the physically inequivalent configurations. The moduli space \mathcal{M} generically has the structure of an orbifold, i.e. it possesses singularities. The singularities of \mathcal{M} appear at the points where the vacuum symmetry group is enhanced or, equivalently, its dimension jumps.

The moduli space \mathcal{M} of the $N = 2$ SYM theory has the natural gauge-invariant vacuum ‘order’ parameter, given by the quadratic Casimir eigenvalue,

$$u \equiv \langle \text{tr} \phi^2 \rangle . \quad (5.1)$$

Eq. (5.1) equally applies to the quantum moduli space, and any gauge group too.

In the $SU(2)$ case, the Higgs field is given by $\phi = \phi^a(x)t^a$, where the $SU(2)$ generators t^a have been introduced, $\text{tr}(t^a t^b) = 2\delta^{ab}$. The classical vacuum configurations satisfying eq. (3.7) can always be put by a gauge transformation into the form $\langle \phi \rangle = \frac{1}{2}at^3$ or, equivalently,

$$\langle \phi \rangle = \frac{1}{2}a\sigma_3 , \quad (5.2)$$

where a complex constant a has been introduced. Hence, semiclassically, one has $u = \frac{1}{2}a^2$ (see sect. III.5 also).

Given a non-vanishing $\langle \phi \rangle$ or $a \neq 0$ semiclassically, the $SU(2)$ gauge symmetry is spontaneously broken to $U(1)$ by the Higgs mechanism. The gauge bosons $W_\mu^\pm = \frac{1}{\sqrt{2}}(A_\mu^1 \pm iA_\mu^2)$ get mass $m = \sqrt{2}a$ from the scalar kinetic term $|\nabla_\mu \phi|^2$,²⁷ whereas the rest of the fields, comprising an abelian $N = 2$ vector multiplet and a scalar one in the t^3 -direction, remain massless. The situation is different when $a = 0$, where the $SU(2)$ symmetry is unbroken, and all the fields are massless. Note that the $SU(2)$ rotations by π , forming the so-called (discrete) *Weyl subgroup* of $SU(2)$, change a to $-a$, so that the corresponding vacuum states are gauge-equivalent. The classical moduli space is therefore given by the upper half of a complex plane punctured at the origin. The semiclassical (weak coupling) region corresponds to the area far away from the origin, while the strong coupling region appears in the vicinity of the origin.

²⁷The corresponding gauginos also get the same mass by supersymmetry, thus forming a massive $N = 2$ vector multiplet.

It should be noticed that, after all quantum fluctuations are taken into account, the *quantum* moduli space \mathcal{M}_q may be very different from the classical one. On the one hand, one should expect on physical grounds that a classical singularity may disappear if the associated massless particle is not stable under quantum corrections. On the other hand, new singularities in the quantum moduli space may appear when a charged particle in the full quantum spectrum of the theory becomes massless which results in the enhanced symmetry of the physical vacuum. Although it is not known how to determine the structure of the quantum moduli space from first principles, it can nevertheless be fixed from a consistency of the full quantum theory (see Part III).

The existence of the quantum moduli space is guaranteed by the non-vanishing *Witten index* [42] and the *non-renormalization theorem* in $N = 2$ supersymmetry [49] (see also ref. [50] and the books [36, 37, 38, 39] for more about the non-renormalization in supersymmetry). As was noticed in sect. 3, the $N = 2$ supersymmetry does not allow a superpotential for the $N = 1$ chiral matter superfields in the $N = 1$ superspace formulation of the $N = 2$ SYM theory. Therefore, the classical flat direction (5.2) remains in the full quantum theory provided that the $N = 2$ supersymmetry is not dynamically broken. A restriction on possible dynamical supersymmetry breaking can be obtained from a calculation of the Witten index $\text{tr}(-1)^F$ which is essentially a topological index counting a difference between the zero-energy bosonic and fermionic states [42]. The supersymmetry is spontaneously broken if the vacuum energy is non-vanishing, which implies the vanishing Witten index. A calculation shows that the Witten index for the $N = 2$ SYM theory is different from zero [42], which means that the $N = 2$ supersymmetry in this theory is not going to be dynamically broken and, hence, the existence of the quantum moduli space is justified.

Though the $SU(2)$ gauge symmetry is spontaneously broken to $U(1)$ in a generic point of the moduli space, the $N = 2$ SYM low-energy effective action is still $N = 2$ supersymmetric. The low-energy effective action is therefore given by an abelian $N = 2$ gauge theory, whose $N = 1$ superspace form is essentially described by eq. (3.16), namely

$$I_{\mathcal{F}}^{\text{abelian}} = \frac{1}{16\pi} \text{Im} \int d^4x \left[\int d^2\theta \mathcal{F}''(\Phi) W^\alpha W_\alpha + \int d^2\theta d^2\bar{\theta} \Phi^\dagger \mathcal{F}'(\Phi) \right] . \quad (5.3)$$

After being written in components, eq. (5.3) yields the kinetic terms

$$I_{\mathcal{F}}^{\text{abelian, kin.}} = \frac{1}{4\pi} \text{Im} \int d^4x \left[-\frac{1}{4} \mathcal{F}''(\phi) F_{\mu\nu} (F^{\mu\nu} - i^* F^{\mu\nu}) + \mathcal{F}''(\phi) |\partial_\mu \phi|^2 - i \mathcal{F}''(\phi) (\lambda \sigma^\mu \partial_\mu \bar{\lambda} - \psi \sigma^\mu \partial_\mu \bar{\psi}) \right] . \quad (5.4)$$

A scalar field theory whose scalar fields are the coordinates of an (internal) manifold is called the *non-linear sigma-model* (NLSM). The NLSM metric G is defined

by the NLSM kinetic terms. In particular, as far as eq. (5.4) is concerned, one has $G_{\phi\phi^\dagger} \sim \text{Im } \mathcal{F}''(\phi)$. If the field ϕ is replaced by its vacuum expectation value a parametrizing the modular space of the $N = 2$ SYM theory, the NLSM metric reduces to the so-called *Zamolodchikov metric* on the moduli space [51],

$$ds^2 = \text{Im } \mathcal{F}''(a) da d\bar{a} = \text{Im } \tau(a) da d\bar{a} , \quad (5.5)$$

where the effective (complexified) coupling constant $\tau(a)$,

$$\tau(a) \equiv \mathcal{F}''(a) , \quad (5.6)$$

has been introduced (*cf.* sect. 3). Unitarity requires the kinetic terms to be positive definite, which implies that

$$\text{Im } \tau(a) > 0 . \quad (5.7)$$

Since \mathcal{F} is a holomorphic function, $\text{Im } \tau$ is a harmonic function and, therefore, it cannot have a minimum on the compactified complex plane. This means that eq. (5.7) *cannot* be satisfied in quantum theory unless the $N = 2$ prepotential \mathcal{F} is not globally defined throughout the moduli space.²⁸ Therefore, to ensure the kinetic terms in the effective action be non-singular, the function \mathcal{F} can only be *locally* defined. It means that we should use different u -coordinates to cover the whole quantum moduli space \mathcal{M}_q , each of them being appropriate only in a certain region of \mathcal{M}_q . It is the structure of singularities on \mathcal{M}_q that tells us how many different local coordinates we really need (Part III).

6 $N = 2$ SYM low-energy effective action and renormalization group

The Zamolodchikov metric is related to the renormalization group and the effective action [51].²⁹ The effective action $\Gamma[\varphi]$ in quantum field theory is defined as the generating functional of *one-particle-irreducible* (1PI) Feynman diagrams. The functional $\Gamma[\varphi]$ is formally given by a Legendre transform of the generating functional $W[\varphi]$ of connected Feynman diagrams. Since the latter has to be renormalized, it introduces a dependence upon the renormalization scale μ into $W[\varphi]$ and $\Gamma[\varphi]$. In spontaneously broken gauge theories, the scale μ is usually identified with the mass scale to be determined by the Higgs mechanism, i.e. the vacuum expectation value

²⁸The only exception is the classical formula (3.15) where τ is a constant.

²⁹See Chapter VIII of ref. [52] for a review.

of the Higgs scalar. The effective coupling constant $e_{\text{eff}}(\mu)$ is defined as the coefficient at the corresponding 1PI vertex function, with its external momenta squared being equal to μ^2 . If a quantum field theory has massless particles, as it usually happens in the gauge theories, one should introduce both an *ultra-violet* (UV) cutoff and an *infra-red* (IR) one, in order to fully regularize the theory. It then becomes important whether momentum integrations in loop diagrams are performed from the UV-cutoff (to be taken to infinity after divergence subtractions) down to zero, or they are only performed down to μ which usually serves as the IR-cutoff. In the latter case, the corresponding effective action $S_W[\varphi; \mu]$ is called the *Wilsonian* effective action [53]. In supersymmetric gauge theories, one should also distinguish between the two definitions of effective action, because of the so-called *Konishi anomaly* [54], which implies that the physical beta-functions to be defined with respect to the two effective actions are also different.³⁰ The Wilsonian effective coupling $e_{\text{eff}}(\mu)$ of a supersymmetric gauge theory is *holomorphically* dependent upon the scale μ , which is not the case for the standard effective action Γ . It is the property that makes the Wilsonian effective action to be preferable in the case of the quantum $N = 2$ SYM theory, whose low-energy effective action has the holomorphic structure due to $N = 2$ supersymmetry. Eqs. (I.5.9), (3.10) and (5.6) imply the following relation between the Zamolodchikov metric and the renormalized (Wilsonian) coupling constants:

$$\text{Re } \tau(\mu) = \frac{\theta(\mu)}{2\pi} , \quad \text{Im } \tau(\mu) = \frac{4\pi}{e^2(\mu)} , \quad (6.1)$$

where the effective vacuum angle (θ -parameter) $\theta(\mu)$ has been introduced. Though being unrenormalized in perturbation theory, the vacuum angle is expected to receive non-perturbative corrections from multi-instanton processes.

Because of the renormalization, the question arises is it the renormalized or the unrenormalized coupling that enters the Dirac quantization condition (I.2.20) and its DZS generalization (I.2.23) ? It does not matter for the $N = 4$ SYM theory which is UV-finite, but it matters for the $N = 2$ SYM theory which is not UV-finite, and, therefore, whose duality properties need to be elaborated further.

The pure (without extra matter) $N = 2$ SYM theory with the gauge group $SU(2)$ is an *asymptotically free* theory. The running of its coupling constant $e(\mu)$ is governed by the beta-function which receives both perturbative *and* non-perturbative (due to instanton corrections) contributions. The perturbative one-loop beta-function can be

³⁰The Konishi anomaly is the field theory analogue of the two-dimensional holomorphic anomaly which is well-known in string theory [55].

calculated by standard perturbation theory, with the result

$$\beta(e) \equiv \mu \frac{de}{d\mu} = -\frac{e^3}{4\pi^2} . \quad (6.2)$$

It is remarkable that the higher-loop orders of perturbation theory do not contribute to that (Wilsonian) beta-function. It can be argued by using either instanton methods [53], or superfield perturbation theory in the ordinary ($N = 1$) covariant superspace [56], in the $N = 2$ extended covariant superspace [57], or in the light-cone $N = 2$ superspace [58]. The extended supersymmetry is crucial in all that approaches. As far as an $N = 1$ supersymmetric gauge theory (with matter) is concerned, the general criterion of perturbative UV-finiteness, based on the knowledge of one-loop beta-function, was given in ref. [59] (see also the book [39]). It should be noticed that all known finite $N = 1$ supersymmetric gauge theories are based on a *simple* gauge group, i.e. they have a single gauge coupling, and their Yukawa couplings are functions of the gauge coupling. Both features are automatic in the extended supersymmetric gauge theories under consideration — see e.g., eq. (3.16).

A simple argument for the absence of all higher loop corrections to the $N = 2$ SYM beta-function (6.2) was given by Seiberg [60]. He noticed that the classical $N = 2$ SYM theory has the global symmetry $SU(2) \otimes U(1)$, where the $SU(2)$ rotates the two spinor superspace coordinates whereas the $U(1)$ (also called *R-symmetry*) multiplies them by a phase: $\theta \rightarrow e^{-i\alpha}\theta$, $\bar{\theta} \rightarrow e^{-i\alpha}\bar{\theta}$ and $\Psi \rightarrow e^{2i\alpha}\Psi$. The *R*-symmetry is anomalous, while the anomaly is given by the index theorem in the presence of an instanton [60],

$$\partial_\mu j_R^\mu = \frac{e^2}{8\pi^2} \varepsilon^{\mu\nu\rho\lambda} F_{\mu\nu} F_{\rho\lambda} , \quad (6.3)$$

which is a non-perturbative phenomenon. The invariance of the perturbative effective action under the $U(1)_R$ symmetry restricts, however, the $N = 2$ prepotential to the form

$$\mathcal{F}_{\text{per}}(\Psi) = \Psi^2 \left[b_1 + b_2 \log \frac{\Psi^2}{\Lambda^2} \right] , \quad (6.4)$$

where b_1 and b_2 are two parameters to be determined from eqs. (3.15) and (6.2), respectively, and Λ is the renormalization-invariant scale at which the gauge coupling becomes strong (see below). Some care should be exercised here, since, though the perturbative effective action is $U(1)_R$ invariant, the effective Lagrangian is actually not. In fact, under an $U(1)_R$ rotation, the perturbative effective Lagrangian, $\mathcal{L}_{\text{per}}^{\text{eff}} = \int d^4\theta \mathcal{F}_{\text{per}} + h.c.$, transforms as

$$\delta \mathcal{L}_{\text{per}}^{\text{eff}} = \frac{\alpha}{4\pi} \varepsilon^{\mu\nu\rho\lambda} \text{tr}(F_{\mu\nu} F_{\rho\lambda}) , \quad (6.5)$$

in agreement with eq. (6.3).

It is clear from eq. (6.4) that the first term represents the classical contribution whereas the second one is a one-loop effect,

$$\mathcal{F}_{\text{per}} = \mathcal{F}_{\text{cl}} + \mathcal{F}_{1\text{-loop}} , \quad (6.6)$$

where $\mathcal{F}_{\text{cl}} = \frac{1}{2}\tau_{\text{cl}}\Psi^2$ and

$$\mathcal{F}_{1\text{-loop}}(\Psi) = \frac{i}{2\pi}\Psi^2 \log \frac{\Psi^2}{\Lambda^2} . \quad (6.7)$$

Therefore, after differentiating eq. (6.6) twice, one finds

$$\frac{4\pi}{e^2(\mu)} + \frac{1}{\pi} \log \frac{a^2}{\mu^2} = \frac{4\pi}{e^2(a)} \equiv \frac{1}{\pi} \log \frac{a^2}{\Lambda^2} , \quad (6.8)$$

where the renormalization-invariant scale Λ is given by

$$\Lambda^2 = \mu^2 \exp \left\{ -\frac{4\pi^2}{e^2(\mu)} \right\} . \quad (6.9)$$

In particular, one easily gets back eq. (6.2).

The effective *field-dependent* coupling constant arises by setting the renormalization scale μ equal to the characteristic scale of the theory given by the vacuum expectation value of the Higgs field: $e_{\text{eff}}(\mu) \rightarrow e_{\text{eff}}(a)$. Eqs. (6.6) and (6.7) imply at $a \rightarrow \infty$ that

$$\tau(a) = \frac{\partial^2 \mathcal{F}_{\text{per}}(a)}{\partial a^2} \sim \frac{i}{\pi} \left(\log \frac{a^2}{\Lambda^2} + 3 \right) . \quad (6.10)$$

The Zamolodchikov metric $\text{Im } \tau(a) \sim \frac{1}{\pi} \log \frac{|a|^2}{\Lambda^2}$ is therefore single-valued and positive in the semiclassical region $u \sim \frac{1}{2}a^2 \rightarrow \infty$, as it should because of unitarity.

Some useful information about multi-valued functions $f(u)$ can be obtained by analyzing their behaviour as u is taken around a closed contour. If there are no special (singular) points inside the contour, the function $f(u)$ will return to its initial value once u has completed the loop. However, if there is a singularity, the multi-valued function $f(u)$ does not usually return to its initial value, which is known as a non-trivial *monodromy*. For example, it follows from eq.(6.10) that the loop around $u \sim \infty$ in the classical moduli space produces a shift $\tau \rightarrow \tau - 2$ because of the branch cut of the logarithm. In its turn, it results in an irrelevant shift of the vacuum angle (τ like \mathcal{F} is also a multi-valued function !). The full story requires knowing the full set of singularities in the quantum moduli space and the monodromy properties of \mathcal{F} (or τ), which are going to be discussed in Part III.

In the IR-region (below Λ), the positivity of $\text{Im } \tau$ is no longer secured by perturbation theory, and the instanton corrections become important. One is left with

an effective abelian gauge theory having vanishing beta-function. In terms of the effective τ -parameter, one has [60]

$$\frac{\theta(a)}{2\pi} + \frac{4\pi i}{e^2(a)} = \frac{4\pi i}{e_0^2} + \frac{i}{\pi} \log \frac{a^2}{\Lambda^2} - \frac{i}{2\pi} \sum_{l=1}^{\infty} c_l \left(\frac{\Lambda^2}{a^2} \right)^{2l}, \quad (6.11)$$

where the infinite sum over the instanton configurations with topological charge l has been introduced. The unknown coefficients c_l can, in principle, be calculated from zero-momentum correlators of the Higgs and gaugino's fields in multi-instanton backgrounds but, in practice, it was only done for a small number of instantons. It is the recent achievement due to Seiberg and Witten [8] who determined the *exact* function \mathcal{F} and, hence, the coefficients c_l altogether (Part III).

According to eq. (6.11), one should expect the full $N = 2$ prepotential to be of the form

$$\mathcal{F}(\Psi) = \frac{1}{2} \tau_{\text{cl}} \Psi^2 + \frac{i}{2\pi} \Psi^2 \log \frac{\Psi^2}{\Lambda^2} + \frac{1}{4\pi i} \Psi^2 \sum_{l=1}^{\infty} c_l \left(\frac{\Lambda^2}{\Psi^2} \right)^{2l}, \quad (6.12)$$

which reproduces eq. (6.11) after differentiating \mathcal{F} twice at $a = \langle \Psi \rangle|_{\theta=0}$.

To conclude this section, as well as the Part II, let me summarize some of the general features, which are apparent in the case of the $N = 2$ SYM theory. Namely,

- the structure of the quantum moduli space does not need to be the same as that of the classical moduli space,
- one should use the Wilsonian effective action to compute the beta-function of renormalization group,
- as far as the (Wilsonian) exact low-energy effective action is concerned, it is the one-loop perturbative effects and non-perturbative instanton contributions that are only relevant, while the perturbation theory beyond one loop is irrelevant.

PART III: Seiberg–Witten theory

In the last Part III of our review, the exact solution to the low-energy effective action in the $SU(2)$ pure (i.e. without $N = 2$ matter) $N = 2$ SYM theory will be described, along the lines of the original work of Seiberg and Witten [8]. Some generalizations to other gauge groups, as well as adding $N = 2$ matter, will also be considered. We conclude with a very short discussion of the impact of that results on confinement and string theory.

1 Quantum moduli space in the $SU(2)$ pure $N = 2$ SYM theory

Unlike the $N = 4$ SYM theory which is supposed to be exactly self-dual in the sense of Montonen-Olive, the $N = 2$ SYM theory cannot be self-dual. It is enough to notice that the ‘fundamental’ fields belong to an $N = 2$ vector multiplet whereas the magnetic monopoles belong to an $N = 2$ scalar multiplet, i.e. an $N = 2$ hypermultiplet (Part II). Nevertheless, the $N = 2$ theory still possesses the *effective* duality, which is now going to be explained.

First of all, one should understand the exact global structure of the quantum moduli space \mathcal{M}_q of vacua. It is entirely determined by *singularities* of \mathcal{M}_q , which should be associated with certain massless physical excitations. Therefore, the global structure of \mathcal{M}_q can be physically motivated. The classical singularity at $u = 0$ is due to extra massless gauge bosons W^\pm , and it results in the gauge symmetry enhancement from $U(1)$ to $SU(2)$. The other singularity at $u = \infty$ ³¹ is due to a branch cut of the logarithm in eq. (II.6.4) which is the one-loop renormalization effect, and it is going to survive in the semiclassical region near $u = \infty$ in the full quantum theory because of asymptotic freedom.

It was postulated by Seiberg and Witten [8] that \mathcal{M}_q has just *two* extra singularities at finite $u = \langle \text{tr} \phi^2 \rangle = \pm \Lambda^2$, where Λ is the dynamically generated quantum scale, while the classical singularity at $u = 0$ in \mathcal{M}_{cl} is absent in \mathcal{M}_q . The absence of a singularity in the origin of \mathcal{M}_q means the absence of massless W^\pm bosons in the full quantum theory. Their presence would otherwise imply a superconformal invariance in the IR-limit, which is not compatible with any scale. Hence, the gauge symmetry is abelian over the whole quantum moduli space, at it never becomes restored to

³¹The moduli space is supposed to be compactified by adding the point at infinity.

the full $SU(2)$ symmetry. The appearance of just two strong coupling singularities, where certain t'Hooft-Polyakov monopoles (or dyons) become massless, is consistent with earlier calculations of the Witten index, $\text{tr}(-1)^F = 2$, and they can be further justified by the ultimate consistency of the solution (see the end of this section). If there were no quantum singularities at all, the coordinate a would be defined globally and unitarity would be lost — see eq. (II.5.7) and the discussion after that.³²

Since the semiclassical masses of the BPS states are protected against quantum corrections (Part II), the BPS mass formula (I.5.17) is valid in the full quantum theory. In terms of the $N = 2$ SYM low-energy effective action, the dual variable a_D is simply given by

$$a_D = \frac{\partial \mathcal{F}(a)}{\partial a}, \quad (1.1)$$

while $\partial a_D / \partial a = \partial^2 \mathcal{F} / \partial a^2 = \tau(a)$.

In physical terms, the a_D is the ‘magnetic dual’ of the ‘electric’ Higgs field a . By $N = 2$ supersymmetry, the a_D has to be a part of the $N = 2$ abelian vector multiplet containing the ‘magnetic dual’ photon A_μ^D . The electro-magnetic duality³³ relates A_μ^D to the ‘fundamental’ gauge potential A_μ . Hence, the magnetic monopoles/dyons couple locally to the dual photon, just like the ‘fundamental’ $N = 2$ hypermultiplets, if present, locally couple to the electro-magnetic gauge potential A_μ . The dual theory looks like the $N = 2$ quantum electrodynamics which is not asymptotically free, and whose ‘magnetic’ beta-function is positive (*cf.* eq. (II.6.2)),

$$\beta_D(e_D) \equiv \mu \frac{de_D}{d\mu} = + \frac{e_D^3}{8\pi^2}. \quad (1.2)$$

The $U(1)$ gauge theory does not contribute to the beta-function (1.2) whose appearance is entirely due to the dual $N = 2$ matter with unit charge coupling to the dual $N = 2$ abelian vector multiplet.

The BPS formula (I.5.17) is also consistent with the appearance of the quantum singularity at $u = +\Lambda^2$ where one should expect $a_D = 0$ but $a \neq 0$. Indeed, a monopole hypermultiplet with charges $n_e = 0$ and $n_m = 1$ would then be massless indeed, in agreement with eq. (I.5.17). Also, since \mathcal{M}_q is supposed to have no singularity at $u = 0$, the semiclassical relation $u \simeq \frac{1}{2}a^2$ cannot be globally valid in the full quantum moduli space.

³²The global \mathbf{Z}_2 symmetry $u \rightarrow -u$ implies that the number of strong coupling singularities must be even. The only fixed points of the \mathbf{Z}_2 symmetry are $u = \infty$ and $u = 0$.

³³An explicit duality transformation will be given in the next section 2.

The effective duality means that the variable $a_D(u)$ should be considered on equal footing with $a(u)$.³⁴ In other words, it does not matter which variable is used to describe the theory — it only depends upon the region (in \mathcal{M}_q) to be described. It is the semiclassical (‘electric’) region (near $u = \infty$) where the preferred local variable is $a(u)$, whereas it is $a_D(u)$ that is the preferred variable near the (‘magnetic’) strong coupling singularity at $u = \Lambda^2$. Also, as was already noticed above, the a_D belongs to the dual gauge multiplet that couples locally to magnetically charged excitations, in the same way that the $a(u)$ locally couples to ‘electric’ excitations. The full theory is of course non-local, which manifests itself in the multi-valuedness of the prepotential \mathcal{F} . In the semiclassical region, the instanton sum in eq. (II.6.12) converges well as long as $a \simeq \sqrt{2u} \rightarrow \infty$. However, the same sum does not make sense outside the convergence domain. Since \mathcal{F} is not an analytic function, the instanton terms in the strong coupling region have to be resummed in terms of some other variables. In particular, near $u = \Lambda^2$, one should expect another (dual) form of the effective Lagrangian,

$$\mathcal{F}_D(\Psi_D) = \frac{1}{2} \tau_{\text{cl}}^D \Psi_D^2 - \frac{i}{4\pi} \Psi_D^2 \log \left[\frac{\Psi_D^2}{\Lambda^2} \right] + \frac{i}{2\pi} \Lambda^2 \sum_{l=1}^{\infty} c_l^D \left(\frac{i\Psi_D}{\Lambda} \right)^l, \quad (1.3)$$

which converges as $\Psi_D \rightarrow 0$. In terms of the original variables, eq. (1.3) describes a strong coupling. The coefficient in front of the logarithm in eq. (1.3) follows from eq. (1.2), and it will be calculated below.

The other singularity at $u = -\Lambda^2$ can be treated in a similar way, after replacing a_D in $\mathcal{F}_D(a_D)$ by $a_D - 2a$ (see below). Hence, three patches are enough to cover the whole moduli space \mathcal{M}_q . Inside of each patch (or *phase*), the theory is weakly coupled in proper variables, and a local effective Lagrangian exists. The relation between the Lagrangians in different phases is however non-local. It is the patching together of the local data about \mathcal{M}_q in a globally consistent way that will completely fix the theory. In other words, it is the absence of a ‘global’ anomaly in the full quantum theory that is important.

Under an $SL(2, \mathbf{Z})$ duality transformation, the section $\begin{pmatrix} a_D(u) \\ a(u) \end{pmatrix}$ on \mathcal{M}_q gets transformed as

$$\begin{pmatrix} a_D(u) \\ a(u) \end{pmatrix} \longrightarrow M \begin{pmatrix} a_D(u) \\ a(u) \end{pmatrix}, \quad (1.4)$$

where $M \in SL(2, \mathbf{Z})$ is nothing but a monodromy matrix, which is entirely determined by the logarithmic terms in eqs. (II.6.12) and (1.3). In particular, in the semiclassical

³⁴We thus confine ourselves to the low-energy effective action, the duality is absent for the full S-matrix !

region near $u = \infty$, one has $u \simeq \frac{1}{2}a^2$ and

$$a_D = \frac{\partial \mathcal{F}(a)}{\partial a} \simeq \frac{i}{\pi} a \left(\log \frac{a^2}{\Lambda^2} + 1 \right) , \quad (1.5)$$

because of asymptotic freedom. Hence, taking the argument u around a loop encircling the point at infinity in \mathcal{M}_q (which looks like \mathcal{M}_{cl} near $u = \infty$) in a clockwise direction ($u \rightarrow e^{2\pi i}u$), one finds that $a \simeq \sqrt{2u} \rightarrow -a$ and ³⁵

$$a_D \rightarrow \frac{i}{\pi}(-a) \left[\log \frac{e^{2\pi i}a^2}{\Lambda^2} + 1 \right] = -a_D + 2a , \quad (1.6)$$

because $u = \infty$ is a branch point of the logarithmic function in eq. (1.5), i.e.

$$\begin{pmatrix} a_D(u) \\ a(u) \end{pmatrix} \longrightarrow M_\infty \begin{pmatrix} a_D(u) \\ a(u) \end{pmatrix} , \quad (1.7)$$

where

$$M_\infty = \begin{pmatrix} -1 & 2 \\ 0 & -1 \end{pmatrix} . \quad (1.8)$$

Near the quantum singularity $u = +\Lambda^2$, the renormalization scale is proportional to $a_D \simeq \langle \Phi_D \rangle \sim 0$, which is the only scale there. In the abelian gauge theory one has $\theta_D = 0$ and, hence, $\tau_D = \frac{4\pi i}{e_D^2(a_D)}$. We can now rewrite eq. (1.2) to the form

$$a_D \frac{d}{da_D} \tau_D = -\frac{i}{\pi} , \quad \text{or} \quad \tau_D = -\frac{i}{\pi} \ln a_D , \quad (1.9)$$

and integrate it further ($\tau_D = -da/da_D$). Hence, near $a_D \sim 0$, one finds in the leading order that

$$a \approx \frac{i}{\pi} a_D \ln a_D . \quad (1.10)$$

It is enough to fix the coefficient in front of the logarithm in eq. (1.3), as well as the monodromy as u goes around the loop encircling $+\Lambda^2$:

$$\begin{pmatrix} a_D(u) \\ a(u) \end{pmatrix} \longrightarrow \begin{pmatrix} a_D(u) \\ a(u) - 2a_D(u) \end{pmatrix} = M_{+\Lambda^2} \begin{pmatrix} a_D(u) \\ a(u) \end{pmatrix} , \quad (1.11)$$

where

$$M_{+\Lambda^2} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} . \quad (1.12)$$

³⁵Eq. (1.6) implies that the mass of the magnetic monopole becomes infinite in the semiclassical limit $a \rightarrow \infty$, as it should (Part I).

The remaning monodromy matrix at $u = -\Lambda^2$ can be calculated from the factorization condition

$$M_\infty = M_{+\Lambda^2} M_{-\Lambda^2} , \quad (1.13)$$

which, in its turn, follows from the fact that a contour around $u = \infty$ can be deformed into two contours, one encircling Λ^2 and another encircling $-\Lambda^2$. One finds

$$M_{-\Lambda^2} = \begin{pmatrix} -1 & 2 \\ -2 & 3 \end{pmatrix} . \quad (1.14)$$

As was already noticed in sect. I.5, a monodromy transformation can also be interpreted as changing the magnetic and electric numbers $q_m = (n_m, n_e)$ by the right multiplication with M^{-1} . The BPS state with vanishing mass, which is responsible for a quantum singularity, should be *invariant* under the monodromy M , i.e. q_m has to be the eigenvector of M^{-1} (or M) with *unit* eigenvalue. It is obviously the case for the magnetic monopole, with $q_m = (1, 0)$ and the monodromy matrix (1.12). Similarly, the eigenvector of $M_{-\Lambda^2}$ in eq. (1.14) with unit eigenvalue is $(n_m, n_e) = (1, -1)$ which is a *dyon* ! ³⁶

In general, (n_m, n_e) is the eigenvector of

$$M_{(n_m, n_e)} = \begin{pmatrix} 1 + 2n_m n_e & +2n_e^2 \\ -2n_m^2 & 1 - 2n_m n_e \end{pmatrix} , \quad (1.15)$$

with unit eigenvalue. The matrix (1.15) would appear as the monodromy matrix for the singularity due to a massless dyon with charges $q_m = (n_m, n_e)$. ³⁷ Again, one finds a consistency with the initial proposal about the existence of only two quantum singularities at $u = \pm\Lambda^2$. Remarkably, no solution to the monodromy factorization condition exist in the case of more (finite number of) strong coupling singularities [10].

For comparison, it should be noticed that the monodromy group generated by the singularities of the *classical* moduli space \mathcal{M}_q is *abelian*, and it reduces to irrelevant shifts of the vacuum angle, $\theta \rightarrow \theta + 2\pi n$, $n \in \mathbf{Z}$.

In conclusion, the general lessons from this section are:

- the classical vacuum degeneracy is not lifted by quantum corrections, even after the non-perturbative instanton contributions are fully taken into account,

³⁶An explicit dyonic solution was constructed by Sen [32].

³⁷The monodromy matrix M_∞ is not of the form (1.15) since it does not correspond to a massless physical state.

- the monodromies around singularities in \mathcal{M}_q represent the duality transformations which either shift the vacuum angle or connect weak and strong coupling,
- the duality is not a symmetry of the theory, though the charges of the *massless states* to be responsible for quantum singularities are invariant under the duality,
- a consistency of the quantum theory severely restricts the global structure of the quantum moduli space \mathcal{M}_q .

2 Duality transformations

The low-energy effective action is given by the $N = 2$ supersymmetric *abelian* gauge theory whose form in $N = 1$ superspace was written down in eq. (II.5.3). Its dual can be explicitly constructed by the Legendre transform, $\mathcal{F}_D(\Phi_D) = \mathcal{F}(\Phi) - \Phi\Phi_D$, where $\Phi_D \equiv \mathcal{F}'(\Phi)$, which implies

$$\mathcal{F}'_D(\Phi_D) = -\Phi. \quad (2.1)$$

The Legendre transform is known to be very similar to a canonical transformation, with $\mathcal{F}'(\Phi)$ playing the role of a canonical momentum. Since the canonical transformations preserve the phase-space measure, it should not be surprising that the Jacobian of the duality transformation is also equal to one.

The second term in eq. (II.5.3) is obviously invariant under the duality transformation,

$$\begin{aligned} \text{Im} \int d^4x d^2\theta d^2\bar{\theta} \Phi^\dagger \mathcal{F}'(\Phi) &= \text{Im} \int d^4x d^2\theta d^2\bar{\theta} (-\mathcal{F}'_D(\Phi_D))^\dagger \Phi_D \\ &= \text{Im} \int d^4x d^2\theta d^2\bar{\theta} \Phi_D^\dagger \mathcal{F}'_D(\Phi_D). \end{aligned} \quad (2.2)$$

As far as the first term in eq. (II.5.3) is concerned, we need a dual W_D^α to the abelian superfield strength W^α . Unlike the duality relation between Φ_D and Φ , the relation between the W_D^α and W^α cannot be local since it includes, in particular, the duality relation between the component (abelian) field strengths $F_D^{\mu\nu}$ and $F^{\mu\nu}$ (see Part I). The component Bianchi identity for the $F^{\mu\nu}$ is a part of the superspace constraint (II.2.17), which is equivalent to

$$\text{Im}(D_\alpha W^\alpha) = 0, \quad (2.3)$$

and it follows from the abelian version of eq. (II.2.13). Hence, the integration over the unconstrained superfield V in the functional integral defining the quantum theory

can be exchanged for the integration over W^α subject to the the constraint (2.3). The latter can be enforced by using a real Lagrange multiplier V_D as follows:

$$\int \mathcal{D}V \exp \left\{ \frac{i}{16\pi} \text{Im} \int d^4x d^2\theta \mathcal{F}''(\Phi) W^\alpha W_\alpha \right\} \simeq \int \mathcal{D}W \mathcal{D}V_D \exp \left\{ \frac{i}{16\pi} \text{Im} \int d^4x \left(\int d^2\theta \mathcal{F}''(\Phi) W^\alpha W_\alpha + \frac{1}{2} \int d^2\theta d^2\bar{\theta} V_D D_\alpha W^\alpha \right) \right\} . \quad (2.4)$$

One finds

$$\int d^4x d^2\theta d^2\bar{\theta} V_D D_\alpha W^\alpha = \int d^4x d^2\theta (\bar{D}^2 D_\alpha V_D) W^\alpha = -4 \int d^4x d^2\theta (W_D)_\alpha W^\alpha , \quad (2.5)$$

after integrating by parts, and using the relations $\bar{D}_\beta \bullet W^\alpha = 0$ and $W_{D\alpha} \equiv -\frac{1}{4} \bar{D}^2 D_\alpha V_D$. The remaining functional integral over W is Gaussian, and it yields the dual action

$$\int \mathcal{D}V_D \exp \left\{ \frac{i}{16\pi} \text{Im} \int d^4x d^2\theta \left(-\frac{1}{\mathcal{F}''(\Phi)} W_D^\alpha W_{D\alpha} \right) \right\} . \quad (2.6)$$

Note that the effective coupling $\tau(a) = \mathcal{F}''(a)$ has been replaced by the dual one, $-1/\tau(a)$, which is nothing but the S-duality (I.5.11). Since

$$\mathcal{F}_D''(\Phi_D) = -\frac{d\Phi}{d\Phi_D} = -\frac{1}{\mathcal{F}''(\Phi)} , \quad (2.7)$$

one finds

$$-\frac{1}{\tau(a)} = \tau_D(a_D) . \quad (2.8)$$

The dual to the whole action (II.5.3) now takes the same form,

$$\frac{1}{16\pi} \text{Im} \int d^4x \left\{ \int d^2\theta \mathcal{F}_D''(\Phi_D) W_D^\alpha W_{D\alpha} + \int d^2\theta d^2\bar{\theta} \Phi_D^\dagger \mathcal{F}'_D(\Phi_D) \right\} , \quad (2.9a)$$

and it can be rewritten as

$$\frac{1}{16\pi} \text{Im} \int d^4x d^2\theta \frac{d\Phi_D}{d\Phi} W^\alpha W_\alpha + \frac{1}{32\pi i} \int d^4x d^2\theta d^2\bar{\theta} (\Phi^\dagger \Phi_D - \Phi_D^\dagger \Phi) . \quad (2.9b)$$

The S-duality (I.5.11) is only a part of the the full duality group (sect. I.5), and it corresponds to the transformation (*cf.* eq. (I.2.6))

$$\begin{pmatrix} \Phi_D \\ \Phi \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \Phi_D \\ \Phi \end{pmatrix} . \quad (2.10)$$

The transformation (2.10) is not a symmetry of the theory, but it relates its two different parametrizations, one being more suitable for weak coupling while the other for strong coupling. It follows from the form (2.9b) of the dual action that there is a symmetry

$$\begin{pmatrix} \Phi_D \\ \Phi \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Phi_D \\ \Phi \end{pmatrix} , \quad \text{where } b \in \mathbf{Z} , \quad (2.11)$$

which only results in an irrelevant shift of the first term in eq. (2.9b) by

$$\frac{b}{16\pi} \text{Im} \int d^4x d^2\theta W^\alpha W_\alpha = -\frac{b}{16\pi} \int d^4x F_{\mu\nu}^* F^{\mu\nu} = -2\pi b n , \quad (2.12)$$

where n is the instanton number (sect. I.3). The transformations (2.10) and (2.11) together generate the full S-duality group $SL(2, \mathbf{Z})$.

Since $a_D(u) = \partial\mathcal{F}(a)/\partial a$, the Zamolodchikov metric (II.5.5) can be rewritten in the explicitly $SL(2, \mathbf{Z})$ -invariant form as

$$ds^2 = \text{Im} (da_D d\bar{a}) = \frac{i}{2} (dad\bar{a}_D - da_D d\bar{a}) = -\frac{i}{2} \varepsilon_{mn} \frac{dv^m}{du} \frac{d\bar{v}^n}{d\bar{u}} du d\bar{u} , \quad (2.13)$$

where the two-dimensional vector

$$v^m \equiv \begin{pmatrix} a_D \\ a \end{pmatrix} \quad (2.14)$$

is considered as a function of u .

3 Seiberg-Witten elliptic curve

A solution to the low-energy effective action or, equivalently, a calculation of multi-valued functions $a_D(u)$ and $a(u)$, was reduced in sect. 1 to the standard *Riemann-Hilbert* (RH) problem of finding the functions with a given monodromy around the singularities. A solution to the RH problem is known to be unique up to a multiplication by an entire function. The last ambiguity can be resolved in our case by the known asymptotical behaviour.

The monodromy matrices (1.12) and (1.14) generate the monodromy group $\Gamma(2)$ which is a subgroup of the modular group $SL(2, \mathbf{Z})$,

$$\Gamma(2) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbf{Z}) , \quad b = 0 \bmod 2 \right\} . \quad (3.1)$$

The fact that the $N = 2$ theory is not self-dual becomes transparent by noticing that the S-duality (I.5.11) having $b = 1$ does not belong to the $\Gamma(2)$. Still, there are other transformations in eq. (3.1) which relate weak and strong coupling, and it is the precise definition of the effective duality in the $N = 2$ theory under consideration. The quantum moduli space is therefore given by

$$\mathcal{M}_q \cong \mathbf{H}^+ / \Gamma(2) , \quad (3.2)$$

where \mathbf{H}^+ is the upper half-plane.

It was the Seiberg-Witten idea [8] to introduce an auxiliary genus-one *Riemann surface* (elliptic curve) whose moduli space is precisely given by \mathcal{M}_q of eq. (3.2), and whose period ‘matrix’ (or elliptic modulus) is precisely the gauge coupling $\tau(u)$. That auxiliary construction automatically guarantees positivity of the Zamolodchikov metric ($\text{Im } \tau > 0$) because of the well known ‘Riemann second relation’ in the theory of Riemann surfaces [62]. In addition, it secures integer monodromy (see below).

The relevant Riemann surface is defined by an algebraic equation

$$y^2(x, u) = (x^2 - u)^2 - \Lambda^4 \equiv \prod_{i=1}^4 (x - e_i(u, \Lambda)) , \quad (3.3)$$

where

$$\begin{aligned} e_1 &= -\sqrt{u + \Lambda^2} , & e_2 &= -\sqrt{u - \Lambda^2} , \\ e_3 &= +\sqrt{u - \Lambda^2} , & e_4 &= +\sqrt{u + \Lambda^2} , \end{aligned} \quad (3.4)$$

and it can be represented in terms of two sheets (complex planes) connected through the cuts $[e_1, e_2]$ and $[e_3, e_4]$. The point at infinity is supposed to be added to each sheet, so that one gets the topology of a torus.

The period ‘matrix’ $\tau(u)$ of the torus is defined by a ratio of its *period* integrals,

$$\tau(u) = \frac{\omega_D(u)}{\omega(u)} , \quad (3.5)$$

where

$$\omega_D(u) = \oint_{\beta} \tilde{\omega} , \quad \omega(u) = \oint_{\alpha} \tilde{\omega} , \quad \text{with } \tilde{\omega} \equiv \frac{dx}{y(x, u)} , \quad (3.6)$$

and (α, β) is a canonical homology basis of the torus.³⁸

Since $\tau = \partial a_D / \partial a$, eq. (3.5) suggests to identify

$$\omega_D(u) = \frac{da_D(u)}{du} , \quad \omega(u) = \frac{da(u)}{du} . \quad (3.7)$$

Hence, both functions $a_D(u)$ and $a(u)$, as well as the prepotential, $\mathcal{F} = \int da a_D(a)$, can be obtained by integration of the torus periods. One finds

$$a_D(u) = \int_{\beta} \lambda , \quad a(u) = \int_{\alpha} \lambda , \quad (3.8)$$

where the meromorphic one-form λ is given by

$$\lambda = x^2 \tilde{\omega} = x^2 \frac{dx}{y(x, u)} . \quad (3.9)$$

³⁸The cycle α can be chosen as a loop around e_1 and e_2 , while the cycle β goes through the cuts and encircles e_2 and e_3 .

The monodromy properties of the periods in eqs. (3.6) and (3.8) around the singularities in \mathcal{M}_q fix them completely. Hence, it remains to identify the singularities, and find the monodromy properties in the case of basis cycles α and β of the Riemann surface (3.3).

The singularities arise when the torus degenerates, which happens if any two of the branch points e_i coincide, i.e. when the *discriminant*

$$\prod_{i < j}^4 (e_i - e_j)^2 = (2\Lambda)^8 (u^2 - \Lambda^4) \quad (3.10)$$

vanishes. It results in the three possibilities:

- (i) $e_2 \rightarrow e_3$ or $u \rightarrow +\Lambda^2$, the cycle $\nu_{+\Lambda^2} \equiv \beta$ degenerates,
- (ii) $e_1 \rightarrow e_4$ or $u \rightarrow -\Lambda^2$, the cycle $\nu_{-\Lambda^2} \equiv \beta - 2\alpha$ degenerates,
- (iii) $e_1 \rightarrow e_2$ and $e_3 \rightarrow e_4$, or $\Lambda^2/u \rightarrow 0$.

Going around a singularity in \mathcal{M}_q results in an exchange of the branch points $e_i(u)$ along certain paths (called *vanishing cycles*) ν shrinking to zero when one of the branch points approaches another one. For example, looping around the singularity $u = +\Lambda^2$ results in the rotation of e_2 and e_3 around each other, so that the cycle α gets transformed to $\alpha - 2\beta$, while the cycle β remains intact. This means that the monodromy action is

$$\begin{pmatrix} \beta \\ \alpha \end{pmatrix} \longrightarrow M_{+\Lambda^2} \begin{pmatrix} \beta \\ \alpha \end{pmatrix}, \quad (3.11)$$

where the monodromy matrix $M_{+\Lambda^2}$ is exactly the one as in eq. (1.12). Similarly, one finds that the monodromy matrix to be derived from the vanishing cycle in the case (ii), near the singularity $u = -\Lambda^2$, is precisely given by the matrix $M_{-\Lambda^2}$ of eq. (1.14). The monodromy M_∞ has to be given by eq. (1.8), just because of the consistency relation (1.13). The approach based on the vanishing cycles is therefore justified. An explicit solution will be given in the next section 4.

The vanishing cycles are closely related to massless BPS states. Given a vanishing cycle ν , it can always be decomposed with respect to the homology basis,

$$\nu = n_m \beta + n_e \alpha, \quad (3.12)$$

where n_m and n_e are integers. One finds at a given singularity that

$$0 = \oint_\nu \lambda = n_m \int_\beta \lambda + n_e \int_\alpha \lambda = n_m a_D + n_e a \equiv Z, \quad (3.13)$$

which corresponds to a massless BPS state with the magnetic and electric charges (n_m, n_e) at the singularity ! Therefore, the dyon charges are just the coordinates

of the corresponding vanishing cycle in the homology basis [4]. Under a canonical change of the homology basis (a duality transformation !), the intersection number

$$\#(\nu^i, \nu^j) = n_m^i n_e^j - n_m^j n_e^i \in \mathbf{Z} , \quad (3.14)$$

has to be invariant. Note that eq. (3.14) is nothing but the DZS quantization condition (I.2.23). Two BPS states are mutually local with respect to each other if eq. (3.14) vanishes, and they are non-local otherwise. There exists the general (Picard-Lefschetz) formula [63] that determines the monodromy for any vanishing cycle (3.12), and it just gives rise to eq. (1.15).

4 Solution to the low-energy effective action

It is not difficult to write down the differential equation for a multi-valued section $(a_D(u), a(u))$ having a given monodromy around known singularities in the moduli space parametrized by a local coordinate u . Consider the *second-order* Schrödinger-type equation in the complex plane u ,

$$\left[-\frac{d^2}{du^2} + V(u) \right] \psi(u) = 0 , \quad (4.1)$$

whose potential $V(u)$ is a meromorphic (single-valued) function with a finite number of poles at some points u_i where, for example, $u_1 = 1$, $u_2 = -1$ and $u_3 = \infty$ as in sect. 1.³⁹ Eq. (4.1) is known to have only two linearly independent solutions, let's call them $a_D(u)$ and $a(u)$. As u goes around any of the poles, there can be a non-trivial monodromy, as in eq. (1.4). As is well known in the theory of differential equations [63], the non-trivial constant monodromies correspond to those poles of the potential that are of second order *at most*.⁴⁰ The general form of the potential in our case is therefore fixed up to a few coefficients,

$$V(u) = \frac{d_1}{(u+1)^2} + \frac{d_2}{(u-1)^2} + \frac{d_3}{(u+1)(u-1)} . \quad (4.2)$$

Eq. (4.1) with the potential (4.2) can be transformed into the standard hypergeometric differential equation, whose explicit solutions are known. It remains to compare its general solution, in terms of a hypergeometric function to be parametrized by the potential residues d_i , with the known asymptotics (sect. 1) at each singularity, in order to identify the coefficients d_i , and hence, fix the particular solutions both for

³⁹We take $\Lambda^2 = 1$ for simplicity.

⁴⁰That singularities are called *regular* [63].

$a_D(u)$ an $a(u)$ in terms of hypergeometric functions [5]. The information contained in the asymptotics is equivalent to that contained in the monodromies (sect. 1).

Having obtained the representation (3.8) for the solution in terms of the auxiliary elliptic curve, one can make a ‘short cut’ by verifying that the right-hand sides of eqs. (3.8) are annihilated by the the second-order differential operator

$$\tilde{\mathcal{L}}(w, \theta_w) = \theta_w \left(\theta_w - \frac{1}{2} \right) - w \left(\theta_w - \frac{1}{4} \right)^2, \quad (4.3)$$

where the new variables $w = u^2$ and $\theta_w \equiv w \partial_w$ have been introduced. Eq. (4.3) defines the hypergeometric system $F(-\frac{1}{4}, -\frac{1}{4}; \frac{1}{2}, w)$. It is easy to check that

$$\partial_u \tilde{\mathcal{L}} = \tilde{\mathcal{L}}_{\text{PF}} \partial_u, \quad (4.4)$$

where another operator

$$\tilde{\mathcal{L}}_{\text{PF}}(w, \theta_w) = \theta_w \left(\theta_w - \frac{1}{2} \right) - w \left(\theta_w + \frac{1}{4} \right)^2 \quad (4.5)$$

has been introduced. In terms of the original variable u , the operator (4.5) takes the form

$$\tilde{\mathcal{L}}_{\text{PF}} = (1 - u^2) \partial_u^2 - 2u \partial_u - \frac{1}{4}, \quad (4.6)$$

while the corresponding differential equation, $\tilde{\mathcal{L}}_{\text{PF}} \psi(u) = 0$, is known as the *Picard-Fuchs* (PF) equation, and it plays the role of eq. (4.1) here. All the periods of the Seiberg-Witten elliptic curve are known to satisfy the PF equation [62, 63]. For those of them, which are given by eqs. (3.6) and (3.7), it was just argued. In our case, matching the asymptotic expansions of the period integrals in accordance with the results of sect. 1 yields the particular combination of hypergeometric functions [5],

$$\begin{aligned} a_D(u) &= \frac{i}{2}(u-1)F\left(\frac{1}{2}, \frac{1}{2}; 2, \frac{1-u}{2}\right), \\ a(u) &= \sqrt{2(u+1)}F\left(-\frac{1}{2}, \frac{1}{2}; 1, \frac{2}{u+1}\right). \end{aligned} \quad (4.7a)$$

Using standard integral representations of the hypergeometric functions [64], one can rewrite eq. (4.7a) to the very explicit form [8],

$$\begin{aligned} a_D(u) &= \frac{\sqrt{2}}{\pi} \int_1^u \frac{dx \sqrt{x-u}}{\sqrt{x^2-1}}, \\ a(u) &= \frac{\sqrt{2}}{\pi} \int_{-1}^1 \frac{dx \sqrt{x-u}}{\sqrt{x^2-1}}. \end{aligned} \quad (4.7b)$$

It is straightforward to calculate the prepotential $\mathcal{F}(a)$ from the explicit expressions given above. For example, one can invert the second equation in eq. (4.7) and

insert the result into the first one, in order to obtain a_D as a function of a . Integrating the latter once with respect to a yields $\mathcal{F}(a)$. For example, actual calculations in the case of large a (the semiclassical region) produce eq. (II.6.12) as expected, now with all concrete values for the instanton coefficients c_l , namely [4]

$$\begin{array}{c|cccccc} l & 1 & 2 & 3 & 4 & 5 & \cdots \\ \hline c_l & \frac{1}{2^5} & \frac{5}{2^{14}} & \frac{3}{2^{18}} & \frac{1469}{2^{31}} & \frac{4471}{5 \cdot 2^{34}} & \cdots \end{array} \quad (4.8)$$

Similarly, one can treat the dual magnetic region near the singularity $u = +\Lambda^2$, where the monopole becomes massless. One finds eq. (1.3) indeed, whose lowest threshold correction coefficients read [4]

$$\begin{array}{c|cccccc} l & 1 & 2 & 3 & 4 & 5 & \cdots \\ \hline c_l^D & 4 & -\frac{3}{4} & \frac{1}{2^4} & \frac{5}{2^9} & \frac{11}{2^{12}} & \cdots \end{array} \quad (4.8)$$

The numbers above were confirmed by multi-instanton calculations [65]. The modular-invariant (uniformizing) coordinate u of \mathcal{M}_q is given by [9]

$$u(a) = \pi i \left(\mathcal{F}(a) - \frac{1}{2} a \partial_a \mathcal{F}(a) \right) . \quad (4.9)$$

- It is the power of *holomorphicity* together with *duality* that determine the whole function \mathcal{F} from its known asymptotics near the singularities.

5 Other groups, and adding $N = 2$ matter

Once the exact low-energy effective action of the $SU(2)$ pure $N = 2$ SYM theory is understood, it is straightforward to generalize the Seiberg-Witten results to other gauge groups [66, 67, 68, 69]. Let us take $G = SU(n)$ for definiteness, where $n = N_c$ is the number of ‘colors’.

The classical moduli space \mathcal{M}_{cl} of the inequivalent vacua is the space of all solutions to eq. (II.3.7) modulo gauge transformations. The vacuum expectation value of the Higgs field can be chosen in the *Cartan subalgebra* (CSA) of G ,⁴¹

$$\phi = \sum_{k=1}^r a_k H_k , \quad \text{where } r = \text{rank } G . \quad (5.1)$$

In the case of $G = SU(n)$, one has $r = n - 1$ and $H_k = E_{k,k} - E_{k+1,k+1}$, where $(E_{k,l})_{ij} = \delta_{ik} \delta_{jl}$. In generic point of \mathcal{M}_{cl} , the gauge group G is spontaneously broken

⁴¹The brackets indicating vacuum expectation values are often omitted in what follows, in order to simplify the formulas.

to $U(1)^r$. When some eigenvalues coincide, a (non-abelian) subgroup $H_P \subset G$ remains unbroken.

The electric charge of the $SU(2)$ theory is replaced by the charge vector \vec{q} belonging to the root lattice $\Lambda_R(G)$ in Dynkin basis of G . The BPS mass formula (without magnetic charges) $m^2(q) = 2|Z_q(a)|^2$, where $Z_q(a) = \vec{q} \cdot \vec{a}$, determines which gauge bosons remain massless for a given background $\vec{a} = \{a_k\}$.

The SCA variables \vec{a} are, however, not invariant under the gauge transformations. They do not even have the residual gauge invariance under the discrete transformations from the *Weyl group* $S(n)$.⁴² The gauge-invariant description is provided in terms of the Weyl-invariant Casimir eigenvalues $u_k(a)$ belonging to $\mathbf{C}^{n-1}/S(n)$. The polynomials $u_k(a)$ parametrizing the CSA modulo the Weyl group can be easily obtained by looking at the characteristic equation

$$\det(x\mathbf{1} - \phi) = 0 , \quad (5.2)$$

whose coefficients are Weyl-invariant. In the case of $SU(n)$, one has

$$\phi = \text{diag}(a_1, a_2, \dots, a_n) , \quad \text{and} \quad \sum_i a_i = 0 . \quad (5.3)$$

Hence, eq. (5.2) yields

$$x^n + x^{n-2} \sum_{i < j} a_i a_j - x^{n-3} \sum_{i < j < k} a_i a_j a_k + \dots + (-1)^n \prod_i a_i = 0 . \quad (5.4)$$

Taking $n = 2$ gives $\phi = \frac{1}{2}a\sigma_3$ and $u \equiv \langle \text{tr}\phi^2 \rangle = \frac{1}{2}a^2$, as expected (sect. II.5). In the case of $SU(3)$, one easily finds

$$x^3 - x \frac{1}{2} \text{tr}\phi^2 - \frac{1}{3} \text{tr}\phi^3 = 0 , \quad (5.5)$$

where

$$\begin{aligned} u &\equiv +\frac{1}{2} \langle \text{tr}\phi^2 \rangle = - \sum_{i < j} a_i a_j = a_1^2 + a_2^2 + a_1 a_2 , \\ v &\equiv -\frac{1}{3} \langle \text{tr}\phi^3 \rangle = - a_1 a_2 a_3 = a_1 a_2 (a_1 + a_2) . \end{aligned} \quad (5.6)$$

Similarly, in the case of $SU(n)$, one finds the *symmetric* polynomials

$$\det(x\mathbf{1} - \phi) = x^n + x^{n-2} c_2(\phi) + \dots + (-1)^j x^{n-j} c_j(\phi) + \dots = 0 , \quad (5.7)$$

where

$$c_j(\phi) = \sum_{n_1 < n_2 < \dots < n_j} a_{n_1} a_{n_2} \dots a_{n_j} . \quad (5.8)$$

⁴²The Weyl group $S(n)$ acts on the weights λ_i of G by permutation.

It is more convenient to introduce linear combinations $Z_{\lambda_i}(a) \equiv \vec{\lambda}_i \cdot \vec{a}$, where $\{\lambda_i\}$ are the weights of the n -dimensional fundamental representation of $SU(n)$. It is the $Z_{\lambda_i}(a)$ that have direct group-theoretical meaning, and that enter the BPS mass formula. The corresponding characteristic equation reads

$$\prod_{i=1}^n (x - Z_{\lambda_i}(a)) = x^n - \sum_{l=0}^{n-2} u_{l+2}(a) x^{n-2-l} \equiv W_{A_{n-1}}(x, u_k) . \quad (5.9)$$

The non-linear transition from $Z_{\lambda_i}(a)$ to $u_k(a)$ is known as a classical *Miura transformation*,

$$u_k(a) = (-1)^{k+1} \sum_{j_1 < j_2 < \dots < j_k} Z_{\lambda_{j_1}}(a) Z_{\lambda_{j_2}}(a) \cdots Z_{\lambda_{j_k}}(a) . \quad (5.10)$$

The polynomial $W_{A_{n-1}}(x, u_k)$ is called the *simple singularity* associated with A_{n-1} (or with $SU(n)$) in the theory of partial differential equations [63], or as the *Landau-Ginzburg* (LG) potential in conformal field theory [52]. In the cases of $SU(2)$ and $SU(3)$, one finds

$$W_{A_1} = x^2 - u , \quad W_{A_2} = x^3 - xu - v . \quad (5.11)$$

Extra massless non-abelian gauge bosons appear in the classical theory whenever

$$Z_{\lambda_i}(a) = Z_{\lambda_j}(a) \quad (5.12)$$

for some $i \neq j$. Eq. (5.12) describes classical singularities which are the fixed points of the Weyl transformations. Hence,

$$\mathcal{M}_{\text{cl}} = \{u_k\} / \Sigma_0 , \quad (5.13)$$

where $\Sigma_0 = \{u_k : \Delta_0(u_k) = 0\}$, and the discriminant [63]

$$\Delta_0(u) = \prod_{i < j}^n (Z_{\lambda_i}(u) - Z_{\lambda_j}(u))^2 = \prod_{\text{positive roots}} Z_{\alpha}^2(u) , \quad (5.14)$$

has been introduced. The discriminant of the simple singularity therefore encodes all information about the classical symmetry breaking patterns in the gauge-invariant way.

The $N = 2$ supersymmetry restricts the form of the low-energy effective action to an $N = 2$ abelian gauge theory with the prepotential \mathcal{F} . The theory contains $r = \text{rank } G$ abelian $N = 2$ vector multiplets which can be decomposed into r $N = 1$ chiral multiplets A_i and r $N = 1$ abelian vector multiplets W_i^α . The $N = 1$ superspace Lagrangian is given by

$$\mathcal{L} = \frac{1}{4\pi} \text{Im} \left[\int d^4\theta \left(\sum_i \frac{\partial \mathcal{F}}{\partial A_i} \bar{A}_i \right) + \int d^2\theta \frac{1}{2} \left(\sum_{i,j} \frac{\partial^2 \mathcal{F}}{\partial A_i \partial A_j} W_i^\alpha W_{\alpha j} \right) \right] . \quad (5.15)$$

Accordingly, the $N = 1$ Kähler potential reads

$$K(A, \bar{A}) = \text{Im} \sum_i \frac{\partial \mathcal{F}(A)}{\partial A_i} \bar{A}_i , \quad (5.16)$$

the effective couplings are

$$\tau_{ij}(A) = \frac{\partial^2 \mathcal{F}(A)}{\partial A^i \partial A^j} , \quad (5.17)$$

and the dual fields are defined by

$$A_{\text{D}}^i = \frac{\partial \mathcal{F}(A)}{\partial A_i} . \quad (5.18)$$

As usual, the leading component of the superfield A_i is called a_i , and similarly for A_{D}^i : $A_{\text{D}}^i|_{\theta=0} = a_{\text{D}}^i$.

Zamolodchikov's metric is defined by

$$ds^2 = \text{Im} \frac{\partial^2 \mathcal{F}(a)}{\partial a_i \partial a_j} da_i d\bar{a}_j , \quad (5.19)$$

where $i, j, \dots = 1, 2, \dots, r$. The metric has to be positively definite,

$$\text{Im} \tau_{ij} > 0 . \quad (5.20)$$

The dual coordinates $a_{\text{D}}^i \equiv \frac{\partial \mathcal{F}}{\partial a_i}$ together with the initial coordinates a_i parametrize a $2r$ -dimensional vector space $X \cong \mathbf{C}^{2r}$. Hence, one arrives at a *vector bundle* which locally looks like $\mathcal{M}_{\text{q}} \otimes X$. The X can be endowed with the *symplectic* form

$$\omega = \frac{i}{2} \sum_i \left(da_i \wedge d\bar{a}_{\text{D}}^i - da_{\text{D}}^i \wedge d\bar{a}_i \right) , \quad (5.21)$$

and the *holomorphic form*

$$\omega_{\text{hol}} = \sum_i da_i \wedge da_{\text{D}}^i . \quad (5.22)$$

We are interested in the sections, $f : \mathcal{M}_{\text{q}} \rightarrow X$, which take the form

$$\begin{pmatrix} a_{\text{D}}^i(u) \\ a_i(u) \end{pmatrix} , \quad (5.23)$$

and are restricted by the condition that the pullback of ω_{hol} vanishes: $f^*(\omega_{\text{hol}}) = 0$.

The Zamolodchikov metric

$$ds^2 = \text{Im} \frac{\partial a_{\text{D}}^i}{\partial u_n} \frac{\partial \bar{a}_i}{\partial \bar{u}_m} du_n d\bar{u}_m \quad (5.24)$$

is invariant under the symplectic transformations $Sp(2r, \mathbf{R})$. In accordance with sect. 4, we should expect that only a subgroup Γ_{M} of the *discrete* group $Sp(2r, \mathbf{Z})$

is going to survive in the quantum theory, the Γ_M being generated by actual monodromies in \mathcal{M}_q . It is also known that the same group $Sp(2r, \mathbf{Z})$ is the modular group of a genus- r Riemann surface, whose generators can be visualized in terms of Dehn twists around homology cycles [62]. Therefore, it is a good idea to look for an auxiliary Seiberg-Witten (SW) curve (a Riemann surface) whose moduli space is precisely given by \mathcal{M}_q . Given the SW curve, the positivity of Zamolodchikov's metric would then be guaranteed. In order to identify the right Riemann surface, one notices that it should have something to do with the simple singularity $W_{A_{n-1}}$ playing the key role in determining the structure of the classical moduli space \mathcal{M}_{cl} . For instance, as is well-known in the two-dimensional $N = 2$ supersymmetric conformal field theory, the classical LG potential is still relevant in determining the structure of the quantum theory [52]. Hence, it is not very surprising that the SW curve exists, and it is given by an algebraic curve [61]

$$y^2 = \left(W_{A_{n-1}}(x, u_k)\right)^2 - \Lambda^{2n} . \quad (5.25)$$

Since eq. (5.25) can be rewritten as

$$y^2 = \left(W_{A_{n-1}} - \Lambda^n\right) \left(W_{A_{n-1}} + \Lambda^n\right) , \quad (5.26)$$

it happens that each classical singularity *splits* into two quantum singularities to be associated with massless dyons, with the distance between them being governed by the quantum scale Λ . Accordingly, every single isolated branch of Σ_0 splits into two branches of Σ_Λ . The points Z_{λ_i} also split,

$$Z_{\lambda_i}(u) \implies Z_{\lambda_i}^\pm(u, \Lambda) , \quad (5.27)$$

and become $2n$ *branch points*. The (SW) Riemann surface itself can be represented as a two-sheeted covering of the Riemann sphere branched at $2n$ points, $Z_{\lambda_i}^+$ and $Z_{\lambda_i}^-$, with cuts running between them. Hence, the SW curve appears to be *hyperelliptic*.

By definition, a Riemann surface is called hyperelliptic, if it admits a meromorphic function with exactly two poles [62]. Then, the ramification (branch) points have branch number 1 and, by the *Riemann-Hurwitz* theorem, the number of branch points is related to the genus h by $2n = 2h + 2$, so that $h = n - 1 = r$.⁴³

A generalization to the other *simply-laced*⁴⁴ Lie groups is now obvious: one should simply replace the simple singularity $W_{A_{n-1}}$ with the proper one, W_{D_n} or W_{E_m} , associated with $SO(2n)$ and $E_{6,7,8}$, respectively.

⁴³In fact, any elliptic curve of genus $h \leq 2$ is hyperelliptic [62].

⁴⁴A simply-laced Lie group has all roots of the same length.

Given a Riemann surface of genus h , there exists h holomorphic abelian differentials ω_k (of the first kind) [62]. As far as the SW curve (5.25) is concerned, they are given by

$$\omega_k = \frac{x^{n-k-1} dx}{y}, \quad k = 1, 2, \dots, n-1. \quad (5.28)$$

The period integrals are defined by

$$A_{ij} = \oint_{\alpha_j} \omega_i, \quad B_{ij} = \oint_{\beta_j} \omega_i, \quad (5.29)$$

while the period matrix is $\tau \equiv A^{-1}B$. Hence, one can identify

$$A_{ij}(u) = \frac{\partial}{\partial u_{i+1}} a_j(u), \quad B_{ij}(u) = \frac{\partial}{\partial u_{i+1}} a_D^j(u), \quad (5.30)$$

similarly to that in eq. (3.7). One finds by integration that [67]

$$a_D^i = \oint_{\beta_i} \lambda, \quad a_i = \oint_{\alpha_i} \lambda, \quad (5.31)$$

where (*cf.* eq. (3.9))

$$\lambda = \frac{\text{const.}}{2\pi i} \left(\frac{\partial}{\partial x} W_{A_{n-1}}(x, u_k) \right) \frac{xdx}{y} \quad (5.32)$$

is an abelian differential of the second kind (with vanishing residues). The constant in eq. (5.32) can be fixed from the known asymptotics of (\vec{a}_D, \vec{a}) .

The quantum charges of the massless dyons associated with quantum singularities are determined by the vanishing cycles (see sect. 3). Indeed, any vanishing cycle ν can be decomposed with respect to a homology basis $(\vec{\alpha}, \vec{\beta})$ on the SW curve,

$$\nu = \vec{q} \cdot \vec{\alpha} + \vec{g} \cdot \vec{\beta}, \quad (5.33)$$

where the charge vector \vec{q} has integer components and belongs to the root lattice $\Lambda_{\mathbf{R}}$, while the charge vector \vec{g} also has integer components but belongs to the dual (simple root) lattice $\Lambda_{\mathbf{R}}^D$. One has (*cf.* eq. (3.13))

$$0 = \oint_{\nu} \lambda = \left(\vec{q} \cdot \oint_{\vec{\alpha}} + \vec{g} \cdot \oint_{\vec{\beta}} \right) \lambda = \vec{q} \cdot \vec{a} + \vec{g} \cdot \vec{a}_D \equiv Z_{(q,g)}, \quad (5.34)$$

where the central charge $Z_{(q,g)}$, entering the BPS mass formula $m^2(q, g) = 2 |Z_{(q,g)}|^2$, appears. Hence, similarly to the $SU(2)$ solution (sect. 3), the quantum numbers can be read off from the vanishing cycles. Since the section (5.23) non-trivially transforms under the duality transformations, the charges $\vec{\nu} = (\vec{g}, \vec{q})$ have to transform accordingly, so that the central charge and the BPS mass remain invariant. The *intersection number*,

$$\nu_i \cap \nu_j \equiv \vec{\nu}_i^T \begin{pmatrix} 0 & \mathbf{1} \\ -\mathbf{1} & 0 \end{pmatrix} \vec{\nu}_j = \vec{g}_i \cdot \vec{q}_j - \vec{g}_j \cdot \vec{q}_i \in \mathbf{Z}, \quad (5.35)$$

is also invariant under a change of homology basis (a duality transformation !), and it yields the generalized DZS quantization condition (*cf.* eq. (I.2.23)). Two BPS states are, therefore, local with respect to each other (i.e. a local Lagrangian containing both particles exists), if and only if the intersection number vanishes.

The rest of calculations is quite similar to the $SU(2)$ case considered in sect. 4. The Picard-Lefschetz formula,

$$M_{(g,q)} = \begin{pmatrix} \mathbf{1} + \vec{q} \otimes \vec{g} & +\vec{q} \otimes \vec{q} \\ -\vec{g} \otimes \vec{g} & \mathbf{1} - \vec{g} \otimes \vec{q} \end{pmatrix} \in Sp(2r, \mathbf{Z}) , \quad (5.36)$$

determines the monodromies from the known charges of a given quantum singularity and vice versa. The period integrals of the SW curve satisfy the (second-order) system of $h = r$ Picard-Fuchs differential equations, and they determine the section (5.23) by eq. (5.30). The information from the semiclassical region provided by the perturbative one-loop beta-function (asymptotic freedom !) fixes the monodromy around infinity or, equivalently, determines the perturbative contribution to the $N = 2$ prepotential (*cf.* eq. (II.6.12)) as

$$\mathcal{F}_{1\text{-loop}}(a) = \frac{i}{4\pi} \sum_{\text{positive roots}} Z_\alpha^2 \log \left[\frac{Z_\alpha^2}{\Lambda^2} \right] , \quad (5.37)$$

where $Z_\alpha(a) = \vec{\alpha} \cdot \vec{a}$ for simply-laced Lie groups. The weakly coupled dual prepotential (in proper dual variables) near a quantum singularity looks like that in eq. (1.3), and it is also fixed by the beta-function of the corresponding abelian $N = 2$ supersymmetric gauge theory (no asymptotic freedom). Putting all together, one arrives at the well-defined Riemann-Hilbert problem, whose unique solution can be calculated by solving the Picard-Fuchs equations subject to the known asymptotics near the singularities. It is then straightforward to calculate the $N = 2$ prepotential \mathcal{F} . For example, in the case of $SU(3)$, the solution can be expressed in terms of the so-called *Appel* functions which generalize the hypergeometric functions to the case of two variables [61, 67].

Let us now briefly discuss what happens when an $N = 2$ matter to be represented by some number (N_f) of $N = 2$ hypermultiplets in the fundamental representation of the gauge group $SU(N_c)$ is added.⁴⁵ Each $N = 2$ hypermultiplet comprises two $N = 1$ chiral superfields $Q(q, \psi_q)$ and $\tilde{Q}(\tilde{q}, \psi_{\tilde{q}})$. Under the internal $SU(2)$ symmetry associated to $N = 2$ supersymmetry, the ‘squarks’ (q, \tilde{q}^\dagger) form a doublet, whereas their ‘quark’ superpartners ψ_q and $\psi_{\tilde{q}}$ are singlets.⁴⁶ The $N = 1$ superpotential in

⁴⁵See e.g., the second paper in ref. [8] and refs. [69, 70] for details.

⁴⁶A ‘mirror’ particle $\psi_{\tilde{q}}$ for each quark ψ_q makes $N = 2$ supersymmetry to be phenomenologically unacceptable. $N = 2$ supersymmetry has to be softly broken to $N = 1$ supersymmetry which, in its turn, is spontaneously broken in realistic models (see the next sect. 6 for an example).

the $N = 2$ abelian gauge theory with matter has some additional terms,

$$V_{\text{matter}} = \sum_{i=1}^{N_f} \left(\sqrt{2} \tilde{Q}_i \Phi Q_i + m_i \tilde{Q}_i Q_i \right) + \text{h.c.} , \quad (5.38)$$

where Φ is the chiral $N = 1$ superfield in the $N = 2$ vector multiplet, and $\{m_i\}$ are mass parameters. Because of eq. (5.38), one should expect both the supercurrents (to be derived from the full action), and the central charges in the supersymmetry algebra (to be derived from the supercurrents) to receive contributions from the matter terms too. Accordingly, the BPS mass formula (I.5.17) gets modified. One finds [8]

$$Z = n_e a + n_m a_D + \sum_k S_k m_k / \sqrt{2} , \quad (5.39)$$

where S_k are the $U(1)$ charges of the matter hypermultiplets. Eq. (5.39) implies that the masses $\{m_i\}$ will enter as the additional parameters in the Seiberg-Witten approach to the low-energy effective action. In particular, the positions of the quantum singularities, as well as the SW curve itself, are all going to be deformed by them.

The R-symmetry anomaly in eq. (II.6.3) is replaced by

$$\partial_\mu j_R^\mu = (2N_c - N_f) \frac{F^* F}{32\pi^2} . \quad (5.40)$$

The perturbative (to all loop-orders) beta-function (II.6.2) is also modified as

$$\beta(e) = \mu \frac{de(\mu)}{d\mu} = - (2N_c - N_f) \frac{e^3}{16\pi^2} , \quad (5.41a)$$

or, equivalently ($\alpha \equiv e^2/4\pi$),

$$\frac{1}{\alpha_{N_f}(\mu)} = \frac{2N_c - N_f}{4\pi} \ln \frac{\mu^2}{\Lambda_{N_f}^2} . \quad (5.41b)$$

In the $SU(2)$ case, eqs. (5.40) and (5.41) tell us that one should take $N_f < 4$, in order to keep the asymptotic freedom. If $N_f = 4$ and there are no ‘quark’ masses, the particular $N = 2$ gauge theory with the $SU(2)$ gauge group and four $N = 2$ matter hypermultiplets is finite to all orders of perturbation theory, and it is expected to be conformally invariant even non-perturbatively. That is obviously consistent with the vanishing R -anomaly (5.40) and the vanishing beta-function (5.41), and it presumably gives yet another example of an exactly self-dual theory in the sense of Montonen-Olive with respect to the S-duality, like the $N = 4$ SYM theory though the details are quite different [8]. There is the ‘flavor’ $SO(8)$ global symmetry in the self-dual $N = 2$ theory with matter, while the related $SO(8)$ triality symmetry is non-trivially mixed with the S-duality (*cf.* the U-duality in a compactified type-II superstring theory, sect. 7).

The global structure of the quantum moduli space and the low-energy effective action crucially depend on the number of ‘flavors’ N_f . In the $SU(2)$ case, if $N_f = 1$, only a (strong coupling) Coulomb phase appears where $\langle\phi\rangle \neq 0$ and $SU(2)$ is broken to $U(1)$, like in the pure ($N_f = 0$) theory considered in the previous sections. If $1 < N_f < 4$, one can have (strong coupling) Higgs phases also, where the gauge symmetry is completely broken while the light scalars parametrize a unique hyper-Kähler manifold (the existence of a hyper-Kähler structure is dictated by $N = 2$ supersymmetry [71]). In the case of general gauge groups with matter, one finds a rich spectrum of vacua having non-abelian Coulomb phases and mixed Coulomb-Higgs phases as well. Many examples, including a construction of the SW curves in the presence of $N = 2$ matter, can be found in the literature [8, 69, 70].

It is remarkable that the choice of an auxiliary manifold (SW curve) is not unique! In fact, it could be any manifold \mathcal{G} whose moduli space is \mathcal{M}_q , and whose period integrals (to be obtained by integration of proper meromorphic forms over \mathcal{G}) coincide with that of the SW curve. For example, a six-dimensional *Calabi-Yau* (CY) manifold is known [72] which is equally good for describing the low-energy effective action of the $SU(3)$ pure $N = 2$ SYM theory like the SW hyperelliptic curve considered above. When the ten-dimensional type-IIB superstring theory is compactified on that CY space \mathcal{G} down to four dimensions, the resulting four-dimensional $N = 2$ supersymmetric string theory contains the $SU(3)$ pure $N = 2$ SYM theory in the point-particle limit $\alpha' \rightarrow 0$. Hence, one should expect generalizations of the Seiberg-Witten duality to string theory, which is another big story (see sect. 7 also).

6 Seiberg-Witten version of confinement

The Seiberg-Witten results about the exact low-energy effective action in the $N = 2$ supersymmetric gauge theories provide some non-perturbative information about the $N = 1$ supersymmetric gauge theories, including the $N = 1$ super-QCD. One should expect, for example, that the quantum moduli in the $SU(2)$ pure $N = 1$ gauge theory are also given by two points $\pm\Lambda^2$ related by a \mathbf{Z}_2 transformation (R-symmetry), because the Witten index is the same for both theories. In that $N = 1$ theory, it is possible to add a mass term $W = m \text{tr}\Phi^2$ to the potential, where Φ is the chiral $N = 1$ superfield (sect. II.3). The mass term lifts the flat direction of the $N = 2$ potential, and it can be considered as a *soft* $N = 2$ supersymmetry breaking term which allows one to define the $N = 1$ SYM theory as the low-energy effective field theory of the $N = 2$ theory. It is believed that the $N = 1$ theory has a mass gap and, hence, a

non-vanishing gaugino condensation vacuum expectation value, $\langle \bar{\lambda}\lambda \rangle \neq 0$ [53]. The existence of the mass gap in the $N = 1$ theory also implies that the dual ‘magnetic’ photon becomes massive by some Higgs mechanism in the vacua corresponding to the two singularities $\pm\Lambda^2$ in the quantum moduli space. The only obvious candidate for the role of the Higgs field is given by the t’Hooft-Polyakov monopole or dyon of the initial $N = 2$ theory. Such ‘dual’ Higgs effect can be interpreted as the dual mechanism to the well-known Meissner effect in the theory of superconductivity, and it can explain quark confinement as the phenomenon arising from the condensation of the magnetic monopoles carrying global quantum numbers.

The relevant terms in the $N = 1$ supersymmetric action with the dual photon and the monopole field read

$$W = m \operatorname{tr} \Phi^2 + a_D M \tilde{M} , \quad (6.1)$$

where M and \tilde{M} are the $N = 1$ chiral superfields representing the monopole, and the second term gives the coupling of the monopole to the dual photon as required by $N = 2$ supersymmetry. Since $a_D = a_D(u)$ and $u = \operatorname{tr} \Phi^2$, one can rewrite eq. (6.1) to the form

$$W(M) = mu(a_D) + a_D M \tilde{M} . \quad (6.2)$$

Vacua correspond to solutions of $dW = 0$, and satisfy $|M| = |\tilde{M}|$, since the latter is necessary for the vanishing of the D -term. Assuming that $du/da_D \neq 0$, one easily finds from eq. (6.2) the equations of motion,

$$m \frac{du}{da_D} + M \tilde{M} = 0 , \quad a_D M = a_D \tilde{M} = 0 . \quad (6.3)$$

Eq. (6.3) has a non-trivial solution: $a_D = 0$ and

$$\langle M \rangle = \langle \tilde{M} \rangle = \sqrt{-m \frac{du}{da_D}} \neq 0 . \quad (6.4)$$

The non-vanishing magnetic order parameter $\langle M \rangle$ implies the mass gap in the $N = 1$ theory by the dual Higgs mechanism, and the confinement of abelian charge as well.

7 Conclusion

In string theory, the Yang-Mills coupling constant is determined by the vacuum expectation value of the *dilaton* field d , while the Yang-Mills vacuum angle is similarly related to the *axion* field ξ in four space-time dimensions,

$$\frac{e^2}{4\pi} = \langle e^d \rangle = \frac{8G}{\alpha'} , \quad \frac{\theta}{2\pi} = \langle \xi \rangle , \quad (7.1)$$

where the string constant α' and the gravitational (Newton's) constant G are both dimensionful. Hence, the S-duality in string theory acts on the complex (dilaton-axion) field $S \equiv \xi + ie^{-d}$, and it is supposed to relate strong and weak couplings. It gives a reason to expect that a strongly coupled string theory may well be represented by yet another weakly-coupled string theory.⁴⁷ The compactified superstrings have another well-established target space duality called *T-duality* [74], which is usually represented by a non-compact discrete group G_T . The T-duality group G_T together with the S-duality group $SL(2, \mathbf{Z})$ are actually the subgroups of an even larger non-compact discrete group G_U known as *U-duality* [75]. The group G_U appears to be a discrete subgroup of the hidden non-compact continuous symmetry known to be present in the extended supergravity theory arising from the compactified superstring theory in the point-particle limit $\alpha \rightarrow 0$ [34]. For example, the $N = 8$ maximally extended supergravity in four spacetime dimensions has a non-compact global symmetry E_7 , while there is an evidence for the existence of a discrete E_7 as the U-duality group G_U in the corresponding (compactified) type-II superstring theory [75].

Also, in the spirit of Seiberg and Witten, it is quite natural to interpret the so-called *conifold* singularities in the moduli space $\mathcal{M}(\mathcal{G})$ of complex structures of a Calabi-Yau manifold \mathcal{G} (in the type-II superstring compactified on that \mathcal{G}) as that coming from the BPS (stable) massless charged hypermultiplets. The latter are usually interpreted as charged massless *black holes* in string theory [76]. The known dual pairs of string theories provide some examples in which the classical moduli in one theory appear as the quantum moduli in the dual one, thus relating \mathcal{M}_c and \mathcal{M}_q in the string theory context. The Seiberg-Witten approach to the extended supersymmetric gauge theories can therefore be further promoted to the level of superstrings in the very natural way. A thorough discussion of the string dualities is however beyond the scope of this paper.

There exists a deep relation between the Seiberg-Witten low-energy effective theory and *integrability* [77, 78]. In particular, the SW solution can be reformulated in terms of certain integrable systems on the moduli space of instantons. The effective dynamics in the space of coupling constants (τ) is governed by the equations belonging to the generalized *KP-Toda hierarchy* whose solutions are known to be naturally parametrized in terms of auxiliary special surfaces, like the SW curves. For instance, the key relations (III.3.8) can be understood as just the action-integrals (in proper parametrization) in the sine-Gordon model [77] !

One can verify that the known prepotentials $\mathcal{F}(a_i)$, $i = 1, 2, \dots, N - 1$, of the

⁴⁷Perhaps, it may be something else than a string theory (M-theory) [73].

N=2 supersymmetric Yang-Mills theory with the $SU(N)$ gauge group satisfy the WDVV-type [79, 80] equations

$$\mathcal{F}_i \mathcal{F}_k^{-1} \mathcal{F}_j = \mathcal{F}_j \mathcal{F}_k^{-1} \mathcal{F}_i, \quad \text{where} \quad (\mathcal{F}_i)_{jk} \equiv \frac{\partial^3 \mathcal{F}}{\partial a_i \partial a_j \partial a_k}. \quad (7.2)$$

The WDVV-equations are known to express the associativity of the algebra of primary fields in (conformal) topological field theory [79, 80]. That observation provides yet another unexpected link between the four-dimensional N=2 SYM low-energy effective action and the two-dimensional topological field theories. The full story is thus far from being closed !

Acknowledgements

I would like to thank J. Fuchs, J. Gomez, K. Konishi, D. Korotkin, W. Lerche, A. Morozov, H. Nicolai, K. Sibold, S. Theisen and C. Vafa for discussions, and A. Bilal, S. J. Gates Jr., M. Matone and P. West for correspondence.

References

- [1] D. Olive, *Exact electromagnetic duality*, University of Wales preprint SWAT/94–95/81, hep-th/9508089.
- [2] L. Alvarez-Gaumé, *Recent developments in supersymmetric $N = 2$ Yang-Mills theory*, Lecture Notes given at CERN in the Fall of 1994, and at the Summer School in High Energy Physics and Cosmology in Trieste, 1995.
- [3] C. Gomez and R. Hernandez, *Electric-magnetic duality and effective field theories*, Madrid preprint FTUAM 96/36, hep-th/9510023.
- [4] W. Lerche, *Notes on $N=2$ supersymmetric gauge theory*, in “Gauge Theories, Applied Supersymmetry and Quantum Gravity”, Proceedings of the Workshop held in Leuven in July 10–14, 1995; Leuven University Press, 1996.
- [5] A. Bilal, *Duality in $N=2$ SUSY $SU(2)$ Yang-Mills theory: a pedagogical introduction to the work of Seiberg and Witten*, Paris preprint LPTENS–95/53, hep-th/9601007.
- [6] J. A. Harvey, *Magnetic monopoles, duality and supersymmetry*, Chicago preprint EFI–96–06, hep-th/9603086.

- [7] P. Di Vecchia, *Duality in supersymmetric gauge theories*, Nordita preprint 96–57–P, hep-th/9608090.
- [8] N. Seiberg and E. Witten, Nucl. Phys. **426** (1994) 19; *Erratum*: **430** (1994) 485; Nucl. Phys. **B431** (1994) 484.
- [9] G. Bonelli, M. Matone and M. Tonin, *Solving $N=2$ SYM by reflection symmetry of quantum vacua*, Padova preprint DPDFD 96/TH/29; hep-th/9610026.
- [10] R. Flume, M. Magro, L. O’Raifeartaigh, I. Sachs and O. Schnetz, *Uniqueness of the Seiberg-Witten effective lagrangian*, Dublin preprint DIAS-STP/96–21; hep-th/9611123.
- [11] R. Rajaraman, *Solitons and Instantons*, North-Holland Publishing Company, 1982.
- [12] T. H. R. Skyrme, Proc. Roy. Soc. **A262** (1961) 237.
- [13] S. Coleman, Phys. Rev. **D11** (1975) 2088.
- [14] S. Mandelstam, Phys. Rev. **D11** (1975) 3026.
- [15] R. F. Dashen, B. Hasslacher and A. Neveu, Phys. Rev. **D11** (1975) 3424.
- [16] L. D. Faddeev and L. Takhtajan, Teor. Mat. Fiz. **21** (1974) 160.
- [17] G. H. Derrick, J. Math. Phys. **5** (1964) 1252.
- [18] T. T. Wu and C. N. Yang, Phys. Rev. **D12** (1975) 3845.
- [19] P. A. M. Dirac, Proc. Roy. Soc. **A133** (1931) 60.
- [20] D. Zwanziger, Phys. Rev. **176** (1968) 1489.
- [21] J. Schwinger, Science **165** (1969) 757.
- [22] G. t’Hooft, Nucl. Phys. **B79** (1974) 276.
- [23] A. M. Polyakov, JETP Lett. **20** (1974) 194.
- [24] P. Goddard and D. I. Olive, Rep. Progr. Phys. **41** (1978) 1357.
- [25] M. K. Prasad and C. M. Sommerfield, Phys. Rev. Lett. **35** (1975) 760.
- [26] E. B. Bogomol’nyi, Sov. J. Nucl. Phys. **24** (1976) 449.

- [27] N. S. Manton, Phys. Lett. **154B** (1985) 397.
- [28] M. F. Atiyah and N. J. Hitchin, *The Geometry and Dynamics of Magnetic Monopoles*, Princeton University Press, 1988.
- [29] B. Julia and A. Zee, Phys. Rev. **D11** (1975) 2227.
- [30] H. Osborn, *Semiclassical methods for quantizing monopole configurations*, in “Monopoles in Quantum Field Theory”, Proceedings of the Monopole Meeting in Trieste, Italy, 1981, eds. N. S. Craigie, P. Goddard and W. Nahm (World Scientific, 1982).
- [31] E. Witten, Phys. Lett. **86B** (1979) 283.
- [32] A. Sen, Phys. Lett. **329B** (1994) 217.
- [33] C. Montonen and D. Olive, Phys. Lett. **72B** (1977) 117.
- [34] A. Font, L. Ibanez, D. Lüst and F. Quevedo, Phys. Lett. **249B** (1990) 35.
- [35] R. Haag, J. T. Lopuszanski and M. Sohnius, Nucl. Phys. **B88** (1975) 257.
- [36] J. Wess and J. Bagger, *Supersymmetry and Supergravity*, Princeton University Press, 1983.
- [37] S. J. Gates Jr., M. T. Grisaru, M. Rocek and W. Siegel, *Superspace: or One Thousand and One Lessons in Supersymmetry*, Benjamin/Cummings Publishing Company, 1983.
- [38] P. West, *Introduction to Supersymmetry and Supergravity*, World Scientific, 1986.
- [39] O. Piguet and K. Sibold, *Renormalized Supersymmetry*, Birkhäuser, 1986.
- [40] R. Grimm, M. Sohnius and J. Wess, Nucl. Phys. **B133** (1978) 275.
- [41] D. Olive and E. Witten, Phys. Lett. **78B** (1978) 97.
- [42] E. Witten, Nucl. Phys. **B202** (1982) 253.
- [43] S. J. Gates Jr., Nucl. Phys. **B238** (1984) 349.
- [44] L. Brink, J. Schwarz and J. Scherk, Nucl. Phys. **121** (1977) 77.
- [45] H. Osborn, Phys. Lett. **83B** (1979) 321.
- [46] M. Sohnius and P. West, Phys. Lett. **100B** (1981) 245.

- [47] S. Mandelstam, Nucl. Phys. **B213** (1983) 149.
- [48] L. Brink, O. Lindgren and B. E. W. Nilsson, Phys. Lett. **123B** (1983) 323.
- [49] N. Seiberg, Phys. Lett. **318B** (1993) 469.
- [50] M. Grisaru, M. Rocek and W. Siegel, Nucl. Phys. **159B** (1979) 429.
- [51] A. B. Zamolodchikov, JETP Lett. **43** (1986) 730.
- [52] S. V. Ketov, *Conformal Field Theory*, World Scientific, 1995.
- [53] V. Novikov, M. Shifman, A. Vainshtein, M. Voloshin and V. Zakharov, Nucl. Phys. **B229** (1983) 394;
V. Novikov, M. Shifman, A. Vainshtein and V. Zakharov, Nucl. Phys. **B229** (1983) 381, 407;
M. Shifman, A. Vainshtein and V. Zakharov, Phys. Lett. **166B** (1986) 329;
M. Shifman and A. Vainshtein, Nucl. Phys. **B277** (1986) 456; *ibid.* **B359** (1991) 571.
- [54] K. Konishi, Phys. Lett. **135B** (1984) 439.
- [55] E. D'Hoker and D. H. Phong, Rev. Mod. Phys. **60** (1988) 917.
- [56] M. Grisaru and W. Siegel, Nucl. Phys. **B201** (1982) 292.
- [57] P. Howe, K. Stelle and P. West, Phys. Lett. **124B** (1983) 55.
- [58] S. V. Ketov, Theor. Math. Phys. **63** (1985) 470.
- [59] L. Lucchesi, O. Piguet and K. Sibold, Phys. Lett. **201B** (1988) 241.
- [60] N. Seiberg, Phys. Lett. **206B** (1988) 75.
- [61] A. Klemm, W. Lerche, S. Yankielowicz and S. Theisen, Phys. Lett. **344B** (1995) 169.
- [62] H. Farkas and I. Kra, *Riemann Surfaces*, Springer-Verlag, 1980.
- [63] V. Arnold, A. Gusein-Zade and A. Varchenko, *Singularities of Differentiable Maps*, Birkhäuser, 1985.
- [64] A. Erdelyi et al, *Higher Transcendental Functions*, McGraw-Hill, 1953.

- [65] N. Dorey, V. V. Khoze and M. P. Mattis, *Multi-instanton check of the relation between the prepotential \mathcal{F} and the modulus u in $N = 2$ SUSY Yang-Mills theory*, Swansea preprint DTP/96/56; hep-th/9606199.
- [66] M. Douglas and S. Shenker, Nucl. Phys. **B447** (1995) 271.
- [67] A. Klemm, W. Lerche and S. Theisen, Int. J. Mod. Phys. **A11** (1996) 1929.
- [68] P. Argyres and A. Faraggi, Phys. Rev. Lett. **74** (1995) 3931.
- [69] U. Danielsson and B. Sundborg, Phys. Lett. **358B** (1995) 273.
- [70] P. Argyres and A. Shapere, Nucl. Phys. **B461** (1996) 437.
- [71] J. Bagger and E. Witten, Nucl. Phys. **B222** (1983) 1.
- [72] S. Kachru, A. Klemm, W. Lerche, P. Mayr and C. Vafa, Nucl. Phys. **B459** (1996) 537.
- [73] E. Witten, Nucl. Phys. **B443** (1995) 85.
- [74] A. Giveon, M. Porrati and E. Rabinovici, Phys. Rep. **244** (1994) 77.
- [75] C. M. Hull and P. K. Townsend, Nucl. Phys. **B438** (1995) 109.
- [76] A. Strominger, Nucl. Phys. **B451** (1995) 96.
- [77] A. Gorsky, I. Krichever, A. Marshakov, A. Mironov and A. Morozov, Phys. Lett. **355B** (1995) 466.
- [78] A. Marshakov, A. Mironov and A. Morozov, *WDVV-like equations in $N=2$ SUSY Yang-Mills theory*, Moscow preprint; hep-th/9607109.
- [79] E. Witten, Surv. Diff. Geom. **1** (1991) 243
- [80] R. Dijkgraaf, E. Verlinde and H. Verlinde, Nucl. Phys. **B352** (1991) 59.